



Published in final edited form as:

Neuroimage. 2021 April 15; 230: 117795. doi:10.1016/j.neuroimage.2021.117795.

Load-dependent relationships between frontal fNIRS activity and performance: A data-driven PLS approach

Kimberly L. Meidenbauer^{a,*}, Kyoung Whan Choe^{a,b}, Carlos Cardenas-Iniguez^a, Theodore J. Huppert^c, Marc G. Berman^{a,d,*}

^aEnvironmental Neuroscience Lab, Department of Psychology, The University of Chicago, 5848 S University Avenue, Chicago, IL 60637, United States

^bMansueto Institute for Urban Innovation, The University of Chicago, United States

^cDepartment of Electrical and Computer Engineering, The University of Pittsburgh, United States

^dGrossman Institute for Neuroscience, Quantitative Biology and Human Behavior, United States

Abstract

Neuroimaging research frequently demonstrates load-dependent activation in prefrontal and parietal cortex during working memory tasks such as the N-back. Most of this work has been conducted in fMRI, but functional near-infrared spectroscopy (fNIRS) is gaining traction as a less invasive and more flexible alternative to measuring cortical hemodynamics. Few fNIRS studies, however, have examined how working memory load-dependent changes in brain hemodynamics relate to performance. The current study employs a newly developed and robust statistical analysis of task-based fNIRS data in a large sample, and demonstrates the utility of data-driven, multivariate analyses to link brain activation and behavior in this modality. Seventy participants completed a standard N-back task with three N-back levels ($N = 1, 2, 3$) while fNIRS data were collected from frontal and parietal cortex. Overall, participants showed reliably greater fronto-parietal activation for the 2-back versus the 1-back task, suggesting fronto-parietal fNIRS measurements are sensitive to differences in cognitive load. The results for 3-back were much less consistent, potentially due to poor behavioral performance in the 3-back task. To address this, a multivariate analysis (behavioral partial least squares, PLS) was conducted to examine the interaction between fNIRS activation and performance at each N-back level. Results of the PLS analysis demonstrated differences in the relationship between accuracy and change in the deoxyhemoglobin fNIRS signal as a function of N-back level in eight mid-frontal channels.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

*Corresponding authors: meidenbauer@uchicago.edu (K.L. Meidenbauer).

Author contribution statement

K.L.M. and K.W.C. formulated the research question, designed the study, and developed the acquisition protocol. K.W.C. implemented the N-Back study. K.L.M. collected the data. K.L.M., C.C.I., and T.J.H. analyzed the data. M.G.B. supervised the project. K.L.M. wrote the first draft. K.W.C., C.C.I., T.J.H., & M.G.B. provided critical revisions. All authors approved the final version of the manuscript for submission.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Data & Code Availability Statement

fNIRS data, accuracy data, analysis code (for Brain AnalyzIR Toolbox, Behavioral PLS, and task accuracy), results, and n-back PsychoPy task code are all publicly available at: <https://osf.io/sh2bf/>

Specifically, greater reductions in deoxyhemoglobin (i.e., more activation) were positively related to performance on the 3-back task, unrelated to accuracy in the 2-back task, and negatively associated with accuracy in the 1-back task. This pattern of results suggests that the metabolic demands correlated with neural activity required for high levels of accuracy vary as a consequence of task difficulty/cognitive load, whereby more automaticity during the 1-back task (less mid-frontal activity) predicted superior performance on this relatively easy task, and successful engagement of this mid-frontal region was required for high accuracy on a more difficult and cognitively demanding 3-back task. In summary, we show that fNIRS activity can track working memory load and can uncover significant associations between brain activity and performance, thus opening the door for this modality to be used in more wide-spread applications.

Keywords

fNIRS; N-back task; Cognitive load; Partial least squares; Working memory; Neural efficiency

1. Introduction

Functional near-infrared spectroscopy (fNIRS) is a neuroimaging modality that has gained traction in recent years due to its versatility to studying brain activity in realistic natural environments. Compared to electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), fNIRS is more robust to motion artifacts and environmental noise, making it an increasingly popular method for studying neural activity outside of standard laboratory experimentation (Pinti et al., 2018; Yücel et al., 2017). fNIRS uses light spectroscopy at near-infrared wavelengths to measure the same cerebral metabolic changes that are measured using functional MRI (Buxton, 2010; Huppert et al., 2006). In both methods, the measurements taken are metabolic proxies for neuronal activity. When neural activity increases, so does the metabolic demand, leading to increased blood flow in the surrounding vasculature. This blood flow causes an increase in concentrations of oxygenated hemoglobin and a decrease in concentrations of deoxygenated hemoglobin (Buxton, 2013; Huppert et al., 2006).

Historically, many of the typical fNIRS study paradigms and analysis techniques mirrored those of task-based fMRI. The vast majority of existing fNIRS studies involved initial data preprocessing (i.e., downsampling, bandpass or wavelet filtering, motion correction), before being converted into oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) concentrations (Huppert et al., 2009; Scholkmann et al., 2014). A general linear model was then applied to create contrasts in HbX changes between task conditions or between task and rest (Cooper et al., 2012; Pinti et al., 2018; Scholkmann et al., 2014). However, it was recently demonstrated that these typical approaches overall fail to account for specific statistical properties of the fNIRS signal, and in doing so, inflate the false positive rate of reported results (Barker et al., 2013; Huppert, 2016).

Due to these issues, new analysis methods have been developed to account for these specific statistical properties of fNIRS. By applying pre-whitening to the linear model to reduce noise correlations and using robust regression to down-weight statistical outliers, these

methods perform better on sensitivity-specificity analyses and show better control of type-I errors (Barker et al., 2013; Huppert, 2016; Santosa et al., 2018). With proper statistical analysis to account for these unique noise properties, fNIRS provides an increasingly rigorous method of neuroimaging. Furthermore, the cost-effectiveness of fNIRS allows for larger sample sizes than in fMRI, lessening the risk of failures in replicability due to small samples (Turner et al., 2018).

One of the more well-studied effects in fMRI and fNIRS is that of cognitive load-dependent changes in frontal and parietal cortical regions (Cui et al., 2011; Fishburn et al., 2014; Herff et al., 2014; Mencarelli et al., 2019; Owen et al., 2005). That is, neural activity in these regions increases with more cognitively taxing and difficult tasks. Typically, this is achieved by increasing the number of items needed to be stored in working memory in a task requiring sustained attention, as in the N-back task (Conway et al., 2005; Kirchner, 1958; Owen et al., 2005). In an N-back task, participants are required to compare a current stimulus to a stimulus presented N items back during continuous presentation. This N may be manipulated (typical values include $N = 0, 1, 2, \text{ or } 3$), thus indicating the number of items in working memory. While some previous fNIRS studies have shown linear increases in frontal activation based on N-back level (Ayaz et al., 2012; Fishburn et al., 2014; Kuruvilla et al., 2013), several others have found non-linear effects (i.e., activation that does not follow the pattern of 3-back > 2-back > 1-back > 0-back; Aghajani et al., 2017; Mandrick et al., 2016). In cases of non-linear increases in activation with greater task demands, researchers have proposed that participants may simply disengage from tasks that are too difficult (Causse et al., 2017). It has also been posited that once participants reach a maximum level of cortical activation during a demanding (but achievable) task, no additional neural “output” can be tapped into to perform well on an even more difficult task (Mandrick et al., 2013).

While these discrepancies are attributed to performance-based limitations on very demanding tasks, not all studies have attempted to explicitly link the load-dependent activation results to task accuracy, and those which did have yielded mixed results. For example, in a study by Causse et al. (2017), correlations were conducted between fNIRS activity (HbO concentration change) and performance at different levels of task difficulty in three tasks: a spatial working memory task, an executive function task, and a flight simulator. To simplify the analyses, the researchers created average activation across experimenter-specified ROIs and collapsed across multiple levels of difficulty in the 3 tasks in order to decrease the number of potential correlations to 18. In that study, they only found one significant correlation between performance and fNIRS activity under high levels of difficulty on the executive function task. A study by Matsuda & Hiraki (2006) examined fNIRS activity in the dorsal PFC while older children and adolescents (7–14 years) played either a puzzle or fighting video game. They also failed to find a correlation between video game performance and HbO concentration change across any of the ROIs created by averaging nearby channels.

In the case of both the Causse et al. (2017) and Matsuda & Hiraki (2006) studies, the lack of significant correlations may have been reflective of an actual absence of link between task difficulty and cortical activity. However, this failure to find an association could also have resulted from methodological/analytic choices by the researchers such as the choice to

employ univariate analyses, which have limitations. Multivariate approaches consider the covariance between dependent variables, in this case channel activity, which differs from univariate analyses that treat each channel as being independent. This has several advantages. First, there can be interesting structure in modeling/examining the covariance between channels in addition to their mean effects. Second, performing mass univariate statistics requires multiple comparison correction, and thus researchers employ various techniques to limit the total number of statistical tests conducted. For these papers, the researchers averaged activity across channels and/or collapsed activity across difficulty conditions. While these are reasonable steps, the end result is that the potential results are constrained by the researcher choices. However, in a multivariate partial least squares (PLS) analysis (McIntosh and Lobaugh, 2004), only one test is conducted to test the significance of the multivariate pattern against a null distribution, so no multiple comparison correction or data averaging is required. A third advantage of a PLS approach is that it is a data-driven method that is less encumbered by particular researcher decisions. In summary, by adopting a multivariate analysis rather than running multiple bi-variate correlations (i.e., fNIRS activity in each condition with performance in each condition), this approach is able to identify whole-brain brain-behavior correlations that vary systematically by task type or difficulty.

Additionally, other work has identified negative relationships between task performance and cortical activation when participants undergo working memory training, resulting in increased neural efficiency (McKendrick et al., 2014). Results of one fNIRS study examining the role of expertise on prefrontal activity and task difficulty suggested that this relationship is a complex one but did not directly link activation to behavior (Bunce et al., 2011). Thus, while there is evidence that performance influences task-evoked fNIRS activity, how exactly this relationship is affected by task demands remains an open question. However, by adopting a data-driven, multivariate approach such as behavioral partial least squares, these relationships can be more robustly investigated.

The primary goal of the current study was to conduct a well-powered validation of the load-dependent fNIRS responses demonstrated in prior work using a traditional verbal N-back task, large sample, and utilizing recently developed robust statistical analytical procedures. It was hypothesized that during the N-back task, prefrontal and parietal cortical activity would be largest for the 3-back task (highest cognitive load), lessened for the 2-back task (medium cognitive load), and smallest for the 1-back task (lowest cognitive load).

A second, exploratory goal was to examine how individual differences in participant accuracy could affect load-dependent fNIRS activity. This work sought to examine whether the relationship between fNIRS activity and performance differed based on task difficulty, specifically by using a data-driven, multivariate partial least squares (PLS) analysis to evaluate this relationship. Behavioral PLS (McIntosh and Lobaugh, 2004) has been used in other neuroimaging modalities (such as fMRI, EEG, and magnetoencephalography (MEG)) as a data-driven approach to extract relationships between neural activity and behaviors of interest (Bialystok et al., 2005; Chang et al., 2017; Krishnan et al., 2011; Lobaugh et al., 2001; McIntosh et al., 2008), but has not yet been implemented in fNIRS research. In the present work, PLS offers a distinct advantage for investigating the complex links between

performance and neural activation over the more standard approach of contrasting high and low performance participants. Specifically, as we are interested in performance and activity at multiple levels of difficulty and not all participants showed a linear load effect on performance (i.e., some participants specifically did poorly on the 3-back but similarly on 1-back and 2-back, others performed similarly on 2-back and 3-back, etc.), the designation of “high” and “low” performers would be somewhat arbitrary. In contrast, the data-driven approach in PLS allows for these associations to be examined without imposing experimenter-specified grouping. As such, the current study also tested the utility of a multivariate PLS approach in fNIRS research to shed light on how the link between performance and neural activation may be affected by task demands.

In summary, the current study was designed to replicate the effect of load-dependent activation in frontal and parietal cortex in fNIRS using more robust statistical analyses. As with several previous studies, we found non-linear activation effects, which are likely attributable to poor performance on the 3-back task. To better elucidate the impact of performance on activation as a function by N-back level, a behavioral PLS analysis was conducted, and provided insight into how the relationship between accuracy and prefrontal activation differs based on task difficulty.

2. Method

2.1. Participants

Seventy adults participated in this study. All participants had normal or corrected-to-normal visual acuity. Participants gave written informed consent before participation and experimental procedures were approved by the University of Chicago’s Institutional Review Board (IRB). Participants were compensated \$26 or 2 units of course credit, plus a performance-based bonus of up to \$10. The full procedure included additional study elements related to a video intervention that were separate from the current work and lasted approximately 15 min. The total duration of the study was between 75 and 90 min.

Two participants were excluded from all data analysis due to participant non-compliance with the study procedures. Six additional participants were excluded from fNIRS analysis due to technical issues (2 participants) or low quality fNIRS data (4 participants), leaving a final sample of 62 participants. Of the 62 participants with usable fNIRS data, 28 were male and 34 were female, and the mean age was 23.6 years ($SD = 6.3$ years).

2.2. fNIRS data acquisition

fNIRS data were collected from a continuous-wave NIRSport2 device (NIRx Medical Technologies, LLC). The wavelengths of emitted light (LED sources) in this system were 760 nm and 850 nm. The data were collected at a sampling rate of 4.5 Hz using the NIRx acquisition software, Aurora fNIRS. The fNIRS cap contained a total of 16 sources and 16 detectors creating 43 total channels covering bilateral frontal cortex (33 channels) and right parietal cortex (10 channels).

2.3. fNIRS optode locations (Montage)

The montage was created using fOLD (fNIRS Optodes' Location Decider; Morais et al., 2018), which allows placement of optodes in the international 10–10 system to maximally cover anatomical regions of interest, as specified by one of 5 parcellation atlases. The AAL2 (Automated Anatomical Labeling, Rolls et al., 2015) parcellation was used to generate the montage, which was designed to provide as much coverage of the prefrontal cortex (PFC) as possible, covering bilateral superior and inferior frontal gyri. This emphasis on frontal cortical areas was decided based on evidence from other N-back studies using fMRI (see Owen et al., 2005 for a meta-analysis) and fNIRS, which have demonstrated that load-dependent changes in HbO and HbR are found across areas of the PFC (Ayaz et al., 2012; Fishburn et al., 2014; Herff et al., 2014; Sato et al., 2013).

The right parietal region was selected as an additional ROI for this task due to evidence that parietal cortical regions are engaged during attention-demanding tasks in fNIRS (Hosseini et al., 2017; Murata et al., 2015). As parietal data quality is usually less consistent than channels unobstructed by hair (such as the forehead), the majority of optodes (12 sources and 12 detectors) were focused on prefrontal regions, leaving only 4 sources and 4 detectors to cover parietal areas. Rather than sparsely covering bilateral parietal cortex, better coverage of right parietal cortex was examined in the current study. Right parietal was chosen as participants would be required to use their right hand to respond (hence activating left motor/sensorimotor areas) during the task and our parietal montage overlapped somewhat with the standard sensorimotor fNIRS montage. As we did not want to have the more anterior channels in our parietal montage to be affected by differences in contralateral sensory or motor-evoked activity (i.e. due to less or more responding based on task difficulty), we opted to focus on right parietal coverage. While verbal working memory storage and rehearsal are more associated with left-lateralized regions of parietal cortex (Awh et al., 1996; Ravizza et al., 2004), some meta-analytic data demonstrate bilateral parietal activation across verbal and non-verbal N-back tasks (Mencarelli et al., 2019; Owen et al., 2005). [Fig. 2]

Gross ROIs from the montage (used in subsequent figures) were defined based on the Brain AnalyzIR Toolbox's depth map function (Santosa et al., 2018). Depth maps show the distance from each fNIRS optode to the superficial cortex of several talairach daemon labeled regions of the Colin27 atlas (Lancaster et al., 2000), which can be used to determine coverage of an ROI based on the montage used. As a topological fNIRS layout cannot access depths greater than approximately 30 mm, the channels (lines) projected over yellow or orange regions in Fig. 3 (representing depths > 30 mm) are ones that do not reach the specified ROI, whereas channels covering green or blue areas are within range of the nearest cortical point within the ROI.

2.4. Procedure

After providing informed consent, experimenters measured the participants' head to determine cap size and placement, then began to set up the cap while participants were taken through task instructions and given an opportunity to practice the N-back task. After the first round of practice, the cap was placed on the participant's head, moving hair as needed to

provide clear access to the scalp for the sources and detectors. Cap alignment was verified based on the international 10–20 location of Cz (Klem et al., 1999). fNIRS data were then calibrated and checked for quality before proceeding. If any channels were not displaying sufficiently high-quality data, placement and hair-clearing were performed again before continuing. Next, participants completed a short round of additional practice (single block of each 1-back, 2-back, and 3-back without trial-by-trial feedback), before continuing to the main round of the N-back task. After completing the experiment, the cap was removed, and participants completed a demographics questionnaire. All experimental procedures were coded and presented using PsychoPy (Peirce et al., 2019).

2.5. N-back task

The experimenter took participants through step-by-step instructions of the N-back task before participants began practice. Participants were told that they would see a sequence of short words that are separated by brief fixations, and that every 2 seconds a word would be presented that should be compared to the word “N” trials back. In the current study, N was 1, 2, or 3. Participants were instructed to press the “m” key every time the current word matched the word N trials back, and to press the “n” key every time the current word did not match the word N trials back [Fig. 1]. Each block began by displaying the N-back level and a fixation cross (5 s). Each task block contained a 15-length pseudorandom sequence of two words, presented for 2 s each for a total of 30 s, followed by 20 s of rest. Therefore, the length of each block was 55 s. To suppress sequence memory formation, the two words used in each block were randomly selected from the eight-word pool (‘WHAT’, ‘HOW’, ‘WHEN’, ‘WHY’, ‘WHERE’, ‘WHO’, ‘THAT’, ‘BUT’), except during the first practice, in which the words “AXE” and “BOX” were used. In addition, the sequence of two words was determined using an m-sequence (base = 2, power = 4; thus one word appeared eight times, and the other word appeared seven times (Buracas and Boynton, 2002; Choe et al., 2014; Choe et al., 2016) to suppress its autocorrelation. In all cases, words were presented in white text on a black background.

It is worth noting that in the current N-back task design, only two stimuli were used in each block, as this would require a button press for each trial. By counter-balancing the correct response using m-sequence so that the number of hits and correct rejections was roughly equal, this design controls for motor-related and/or event-related global hemodynamic responses. That is, it minimizes the effects of brain activity due to slow ‘odd’ events such as pressing the key that corresponds to a hit when, for example, only 20% of trials include hits (Choe et al., 2014; Jack et al., 2006; Sirotnin et al., 2012) and thus could confound the signals of interest (i.e., the amplitude of n-back block signal). Additionally, this task design using m-sequence counter-balancing of two-stimuli is less prone to two well-known characteristics of the N-back task that complicate the reporting and interpretation of its results, including dissociable effects of different types of errors (i.e., false alarms and misses; Meule, 2017; Oberauer, 2005) and the lure effect (i.e., lures eliciting more errors; Kane et al., 2007; Moore and Ross, 1963; Oberauer, 2005). An m-sequence of length of 15 ($2^4 - 1$) contains all permutations of length 4 by design (Buracas and Boynton, 2002), so it can automatically eliminate lures and the need for dissociating false alarms and misses.

After the experimenter took participants through the N-back instructions participants performed the first round of N-back practice, consisting of 9 blocks. In this first practice, accuracy feedback was provided on a trial-by-trial level as well as at the end of each block. Participants completed 3 blocks of 1-back, then 3 blocks of 2-back, and then 3 blocks of 3-back. After the fNIRS cap was set up, participants began the second round of practice designed to mimic the conditions of the real task more closely. In this practice, participants performed a single block of 1-back, then 2-back, then 3-back, without trial-by-trial feedback. The main N-back task involved 18 blocks, with 6 blocks of each N-back level, pseudorandomly presented.

Participants received a performance-based bonus during this round of N-back task, wherein performance > 90% on a block earned an additional 40 cents per block, > 80% earned an additional 30 cents per block, and > 60% earned an additional 20 cents per block. Performance under 60% did not yield a cash bonus in this study. Participants were informed of their performance on each block and total bonus directly following the 30 s of task.

2.6. Analysis

2.6.1. Behavioral analysis—Accuracy on the N-back task was calculated by taking the average accuracy over the 6 blocks of each N-back level. This accuracy measure was based on the number of correct responses out of the total number of trials which required a response (i.e., 14 trials required responses on 1-back, 13 trials on 2-back, etc.), and therefore, incorrect responses included misses and wrong key presses. Overall, rates of non-responses were relatively low. Across the 6 blocks of each n-back level, participants failed to make a response on an average of 2.29 trials ($SD = 1.56$) out of 84 total for the 1-back task, an average of 2.6 trials ($SD = 1.92$) out of 78 total for the 2-back task, and an average of 2.6 trials ($SD = 1.82$) out of 72 total for the 3-back task. Normalized across total trials per condition, on average participants failed to respond in 2.7% of 1-back trials, 3.3% of 2-back trials, and 3.6% of 3-back trials. Reaction time (RT) was examined both across all responses and for correct responses only, and average RT for each N-back level was also calculated by averaging across the 6 blocks. Statistical analysis was conducted using R version 3.5.1 (R Core Team, 2018). Accuracy- and RT-level differences between levels of the main N-back task, were analyzed with a repeated measures ANOVA using function ‘ezANOVA’ in the ‘ez’ package (Lawrence, 2016), effect size was calculated using the function ‘eta_sq’ in the ‘sjstats’ package (Lüdtke, 2020), and Bonferroni corrected post-hoc contrasts were conducted using paired t-tests in the R ‘stats’ package.

2.6.2. fNIRS data analysis: quality check—fNIRS data were first loaded into the HOMER2 software package (Huppert et al., 2009) for visual inspection and segmentation of the main N-back trials from practice trials. Visual inspection was done to examine overall data quality (at the level of the participant) and to assess the quality of the parietal data, which was much noisier and variable than the frontal data. Visual inspection was performed by examining the power spectral density plots for all channels to identify the presence of a cardiac oscillation, which is typically around 1 Hz (Tong et al., 2011). The presence of this cardiac signal is a good indicator that the optical density signals are successfully coupled with a physiological hemodynamic response (Hocke et al., 2018). This method was used to

do a first pass evaluation. Based on this visual inspection, 4 participants with unusable data (defined as 5 or fewer clean channels) were identified and excluded from further analyses. Parietal data quality was also examined and logged to determine whether analysis of this region would be fruitful. Of the 62 participants that were kept, 17 had fully usable parietal data, 20 had mostly usable parietal data (at least half of channels showing good physiological coupling), and 25 had unusable parietal data (only a few usable channels or none). It should be noted that, as the statistical analysis downweights noisy channels in the linear model (see next section), including these channels will not increase the likelihood of a false positive effect, but the power to detect an effect in this area is reduced.

2.6.3. fNIRS data analysis: pre-processing pipeline and task-based activation

—fNIRS data were then analyzed using the NIRS Brain AnalyzIR Toolbox (Santosa et al., 2018). Using this toolbox, the .nirs data (raw light intensity) were loaded into the program, converted into optical density, then converted to oxygenated (HbO) and deoxygenated (HbR) hemoglobin concentrations using the modified Beer-Lambert law (Strangman et al., 2003).

Once the data were in the form of HbO and HbR concentrations, first level (subject-level) statistics were calculated. As alluded to previously, fNIRS data have unique statistical properties that are not accounted for by typical fMRI-based analysis and can inflate the type-I error rate (Huppert, 2016). In particular, unlike fMRI, fNIRS suffers from serially correlated errors (due to a higher sampling rate than the physiological signal of interest) and heavy-tailed noise distributions (due to motion-related artifacts and often, large differences in SNR between channels and between participants; Huppert, 2016). To correct for these issues, the first level general linear model run on individual participants' data uses an autoregressive, iteratively reweighted least-squares model (AR-IRLS). The AR-IRLS model employs an auto-regressive filter (pre-whitening) to deal with the serially correlated errors and uses robust weighted regression to iteratively down-weight outliers due to motion artifacts (Barker et al., 2013). This model saves both the subject level regression coefficients and their error-covariance matrices to be used in statistical tests and contrasts for each subject, and eventually, for use in second level (group-level) analyses.

Based on research investigating the sensitivity and specificity of basis sets in fNIRS as a function of signal quality and task period (Santosa et al., 2019), a canonical HRF basis set was selected for this analysis. Work by Santosa et al. (2019) found that for tasks of sufficiently long durations (>10 s, as in the current study), the canonical HRF performs best in a sensitivity-specificity (ROC) analysis. The canonical model has lower degrees of freedom than a full deconvolution of the raw hemodynamic response (finite impulse response, or FIR model), which improves performance on ROC analysis. This is true at durations of more than 10 s, even though there may be a mismatch between the shape of the canonical HRF and the actual hemodynamic response (Santosa et al., 2019).

Based on the output of the first level statistical models, 3 subjects with undue leverage for the group analyses were calculated (those which contribute significant leverage towards the group results, defined by subject-level leverage of $p < 0.05$) and were removed from group-level analyses. Next, second-level statistical models were calculated, which use the full covariance from the first-level models to perform a weighted least-squares regression

(Santosa et al., 2018). Robust regression was also applied to the second-level model to down-weight outliers at the group-level. The results of this analysis were used for group-level contrasts between N-back levels at each channel.

Group activation results are reported as statistical maps using Benjamini-Hochberg false-discovery rate-corrected p -values (e.g., q -values; Benjamini and Hochberg, 1995). This FDR correction was applied to all the data in the second-level analysis, including 43 channels, oxy- and deoxy-hemoglobin, and 3 conditions, making the correction very conservative over all tests.

2.6.4. fNIRS data: behavioral PLS analyses—Behavioral PLS analysis (Berman et al., 2014; McIntosh and Lobaugh, 2004; Krishnan et al., 2011; <https://www.rotman-baycrest.on.ca>) was conducted to identify significant relationships between fNIRS activity and task performance as a function of N-back level. PLS (partial least squares) analysis is a multivariate, data-driven approach often used to examine brain-behavior associations in neuroimaging research by relating two sets of data to one another (Krishnan et al., 2011), in our data these two sets were channel activity and N-back level/condition. Because this was a behavioral PLS and not a task PLS, we were interested in how the correlation between performance and channel activity varied by n-back level. In this study, the fNIRS brain data consisted of the regression coefficients (β) from the first-level statistical model (AR-IRLS), corresponding to changes in HbO or HbR for each N-back level relative to baseline for each participant. The behavioral data consisted of average accuracy for each N-back level across blocks of the main N-back task for each participant. Thus, for each behavioral-PLS (HbO or HbR) analysis, each participant had 129 brain activity values (activation betas for each of three N-back levels for 43 channels) and three behavioral values i.e., average accuracy for each of three N-back levels.

Like other forms of PLS, behavioral PLS relies on the singular value decomposition (SVD) of a covariance matrix. For matrix X of brain activity (i.e., fNIRS activation betas) and Y of behavioral data (accuracy), X and Y were first independently mean centered and normalized within N-back condition. The brain data were 62×43 , i.e., 62 subjects \times 43 channels, and there were 3 of these matrices (one for each load). The behavioral matrix Y was a 62×1 vector of average accuracy for that load level and there were 3 of these vectors (one for each load level). Then, the product between vector Y and matrix X was calculated, i.e., $Y' * X$, which yields a vector of correlations between X and Y of size 1×43 . There are 3 such vectors, one for each n-back level. These vectors are then stacked on top of each other, yielding a 3×43 cross-product matrix R . This cross-product matrix R is the input for the SVD, which decomposes R into three matrices: $R = U \Lambda V^T$, where the singular vector U represents the decomposition of R in behavioral/condition space, the singular vector V represents the decomposition of R in neural activity space, and Λ is the diagonal matrix of singular values, that quantifies the weighting of each of the singular vectors. As such, the behavioral saliences (U) represent the N-back level-dependent differences in the brain-behavior correlation between accuracy and channel activity, and the brain saliences (V) represent the fNIRS channel-dependent differences in this brain-behavior correlation by n-back level.

The goal of this analysis was to find the linear combination of conditions and brain activity that maximized their covariance. These weighted patterns are calculated from the saliences U and V and are referred to as latent variables (LVs). The LVs maximize the covariance between brain activity and accuracy as a function of N-back level. Ten thousand permutation tests were performed to obtain p -values for each latent variable (LV) and 10,000 bootstrapped samples with replacement were created to generate the 95% confidence intervals for the mean correlation between fNIRS activity and performance by condition for each channel. The bootstrap ratios (saliency[weights]/SE[reliability]) measure the reliability of the brain-behavior relationship at each channel, and a larger bootstrap ratio indicates a stronger and/or more consistent contribution to the LV. In this study, channels with bootstrap ratios larger than +2 or smaller than -2 were determined to be statistically significant as these bootstrap ratios can be interpreted as z-scores.

Before running the PLS analysis, histograms of the fNIRS beta values were plotted to examine whether the brain data contained any extreme outliers that may bias the PLS analysis and would ordinarily be removed in the AnalyzIR Toolbox's robust regression (Huppert, 2016). One participant contained extreme outliers at channel 29 (i.e., beta values < -100 and > 100), and was therefore excluded from PLS analysis. However, the direction and significance of results did not change if this participant was included.

2.7. Data & code availability

Data, analysis code, results, and experiment code are publicly available at: <https://osf.io/sh2bf/>.

3. Results

3.1. Behavioral results: N-back performance

Results of the repeated measures ANOVA examining accuracy as a function of N-back level in the main task yielded a significant effect of N-back level on accuracy, $F(2,122) = 85.1$, $p < 0.001$, $\eta_p^2 = 0.58$, 95% CI [0.47, 0.66]. As expected, accuracy for the 1-back task ($M = 0.90$, $SD = 0.11$, 95% CI [0.88, 0.93]) was significantly better than accuracy for the 2-back task ($M = 0.78$, $SD = 0.17$, 95% CI [0.74, 0.82], $p < 0.001$) and for the 3-back task ($M = 0.72$, $SD = 0.16$, 95% CI [0.68, 0.76], $p < 0.001$). Accuracy for the 2-back task was also significantly higher than for the 3-back task ($p < 0.001$). [Fig. 4]

Results for the repeated measures ANOVA examining reaction time (RT) across all responses by N-back level yielded a significant effect of N-back level on reaction time, $F(2,122) = 50.1$, $p < 0.001$, $\eta_p^2 = 0.45$, 95% CI [0.32, 0.55]. Pairwise comparisons demonstrated that RT for the 1-back task ($M = 0.69$, $SD = 0.17$, 95% CI [0.65, 0.74]) was significantly faster than RT for the 2-back task ($M = 0.81$, $SD = 0.19$, 95% CI [0.76, 0.86], $p < 0.001$) and the 3-back task ($M = 0.81$, $SD = 0.19$, 95% CI [0.76, 0.85], $p < 0.001$). However, reaction times for the 2-back and 3-back task were not significantly different ($p = 1$) [Fig. 4]. When analyses were conducted on RTs for correct responses only, the results were the same (overall effect of N-back level ($p < 0.001$) driven by faster RTs on 1-back

relative to 2-back ($p < 0.001$) or 3-back ($p < 0.001$) with nearly identical RT on 2-back and 3-back ($p = 1$).

3.2. fNIRS activation results

3.2.1. Activation vs. baseline in main N-back task—In the GLM, baseline is defined by everything that is not a task event (i.e., is a test against the DC regressor in the model). Relative to baseline, significant increases in oxygenated hemoglobin (HbO) were found for 1 channel (medial superior frontal gyrus) for the 1-back task and for 5 frontal channels and 1 parietal channel for the 2-back task. No channels showed significant increases in HbO concentrations for the 3-back task. [Table 1] No channels showed significant changes in HbR for any N-back level. Group-level activation as the HRF time series for each of the N-back levels across 5 ROIs is plotted in Fig. 5.

3.2.2. fNIRS contrasts: 2-back vs. 1-back—For HbO, 18 channels in the bilateral frontal and right parietal cortex showed significantly larger ($q < 0.05$) increases during the 2-back task relative to the 1-back task. No channels yielded larger HbO increases for the 1-back task relative to 2-back.

For HbR, 7 channels, primarily in bilateral inferior frontal gyrus (IFG), displayed larger decreases for 2-back over 1-back. Additionally, 5 channels, primarily in the right middle occipital gyrus, yielded larger decreases in HbR for 1-back relative to 2-back. [Fig. 6, Top Panel]

3.2.3. fNIRS contrasts: 3-back vs. 1-back—For HbO, 8 channels, primarily in left and right IFG, yielded significantly larger increases for 3-back relative to 1-back. Larger HbO increases for 1-back over 3-back were found in 7 channels, primarily located in the right inferior parietal cortex and left superior frontal gyrus (SFG).

For HbR, 5 channels (4 prefrontal, 1 inferior parietal), demonstrated larger deactivation in the 3-back task compared to 1-back. Eight channels (4 frontal and 4 occipito-parietal) showed the opposite pattern. [Fig. 6, Middle Panel]

3.2.4. fNIRS contrasts: 3-back vs. 2-back—For HbO, 22 channels showed significantly larger increases during the 2-back task compared to the 3-back task. These channels covered bilateral frontal and right parietal areas. Only one frontal channel was greater for the 3-back relative to the 2-back task.

For HbR, 9 channels distributed across bilateral frontal and right parietal cortex showed larger decreases for the 2-back task relative to the 3-back, and 4 channels (2 in medial SFG and 2 in inferior parietal cortex) showed the inverse pattern. [Fig. 6, Bottom Panel]

3.2.5. fNIRS contrasts summary—Group level activation maps and contrasts between N-back conditions showed the most consistent results in the 2-back task relative to baseline and comparing activity during the 2-back task relative to the 1-back task. The consistently higher HbO and lower HbR concentration changes during the 2-back task, but not 3-back task, suggest that a minimum level of accuracy may be needed to elicit reliable activation in

the fronto-parietal cortical regions examined. By minimum accuracy, we mean that if participants are not performing with a high enough accuracy they may not actually be engaged in the task because it has become too difficult. Participants overall performed more poorly on the main 3-back task. For the 59 participants used in group-level analysis of the main N-back task,¹ the average accuracy was 73.6% for the 3-back task. In comparison, average accuracy for these 59 participants was 80% for the 2-back task and 92.3% for the 1-back task.

3.3. Behavioral PLS analysis: fNIRS activity ~ task performance

Separate behavioral PLS analyses were run to relate performance to concentration changes in HbO and HbR by condition (i.e., 1-back, 2-back, and 3-back). Though no statistically significant LVs were found for oxyhemoglobin (HbO), the first latent variable from the analysis with deoxyhemoglobin concentrations (HbR) was significant and explained 51% of the crossblock covariance ($p = 0.025$). LVs 2 and 3 in this analysis were not significant (all $p > 0.9$). For the significant LV 1, eight superior frontal gyrus (SFG) channels (#1, #3, #4, #8, #9, #12, #16, and #25) showed N-back level dependent changes in the relationship between HbR and task accuracy [Table 2]. All of these significant SFG channels had bootstrap ratios > 2 , indicating the direction of the brain-behavior relationship was the same across all eight channels. In addition, one channel in the left inferior frontal gyrus had a bootstrap ratio in the opposite direction (BSR = -2.4). For the eight superior frontal gyrus channels, a larger reduction in HbR (equivalent to increased neural activity) was positively correlated with higher performance on the 3-back task, unrelated to activity on the 2-back task, and negatively correlated with performance on the 1-back task. In summary, this suggests that the metabolic demands placed on the prefrontal cortex that are necessary to achieve a high level of accuracy varies as a consequence of how difficult the task is. [Fig. 7]

4. Discussion

The initial, confirmatory aim of this study was to further validate the use of fNIRS for measuring cognitive load with a large sample and utilizing recently developed robust statistical tools. Though a number of previous fNIRS studies have examined prefrontal activity using attention demanding working memory tasks such as the N-back (Aghajani et al., 2017; Ayaz et al., 2012; Fishburn et al., 2014; Kuruvilla et al., 2013; Sato et al., 2013), recent work has demonstrated that due to the unique statistical properties of fNIRS, the standard analysis approach (based on fMRI) can severely inflate the false positive rate (Huppert, 2016). In addition, discrepancies between studies that may be related to task performance have been demonstrated across several studies. Therefore, to provide convergent evidence for previous studies examining load-dependent changes in PFC and parietal cortex, a standard N-back task was employed with a large sample of participants, and using the recently developed Brain AnalyzIR Toolbox (Santosa et al., 2018) to deal with these fNIRS-specific statistical properties. As an exploratory aim, a behavioral PLS analysis was conducted to directly examine how performance and activation might be dependent

¹Three of the 62 usable fNIRS participants (#s P42, P67, and P70) were removed due to undue group-level leverage, see section 2.6.3. fNIRS Data Analysis: Pre-processing Pipeline and Task-Based Activation.

upon cognitive load, which may explain non-linear load effects in this study and in previous work.

Overall, the fNIRS results were consistent with the general hypothesis that tasks placing higher demands on attention and working memory would lead to increased frontal and parietal activation as measured by HbO and HbR concentration changes. This was most evident by the widespread frontoparietal activation elicited by the 2-back task relative to the 1-back task.

However, activity in the 3-back task did not follow the hypothesized pattern. As mentioned previously, this non-linear load effect has been demonstrated in other fNIRS studies (Aghajani et al., 2017; Mandrick et al., 2013; Mandrick et al., 2016), and seems likely due to poor performance on the task. These results are consistent with the idea that when task demands exceed the current mental capacity of participants, they may disengage from the task and potentially, fail to recruit the necessary cognitive resources (Mandrick et al., 2013).

Interestingly, results of the PLS analysis that incorporated individuals' accuracy by N-back level demonstrated evidence for an interaction of load and performance in the recruitment of the PFC. Specifically, this multivariate approach showed that changes in deoxyhemoglobin concentrations (HbR) in the medial SFG, which were only uncovered when examining the relationship between brain-behavior, but not when behavioral performance was not accounted for. Here, greater reductions in HbR (i.e., more activation) was positively related to performance on the 3-back task, unrelated to accuracy in the 2-back task, and negatively related to accuracy in the 1-back task. This pattern of results suggests that more automaticity during the 1-back task (i.e., less activation) led to better performance on this relatively easy task, and extensive recruitment of the PFC was required for high accuracy on the more difficult, cognitively demanding 3-back task.

This effect may reflect what has been proposed by the neural efficiency hypothesis: that participants with overall greater cognitive processing ability will show less activation during easy tasks and more during difficult tasks (Dunst et al., 2014; Neubauer and Fink, 2009). This is thought to result from the lower metabolic demands that a "more efficient" brain requires during cognitive tasks. Though the neural efficiency hypothesis is often framed as reflecting individual differences in intelligence, there is also evidence that this effect occurs as a result of more efficient strategies after adequate practice on a specific task (Sayala et al., 2006). Thus, one possibility for this interaction of task difficulty and prefrontal activation is that this reflects individual differences in the learning and adoption of effective strategies during practice. Interestingly, recent neuroimaging work has shown that individuals whose brains are in a more scale-free or fractal state tend to reap the benefits of practice to a greater degree than do those starting in a less scale-free state (Kardan et al., 2020). Though scale-free neural dynamics have been demonstrated in fMRI and EEG (Churchill et al., 2016; Kardan et al., 2020), whether this signal can be extracted from fNIRS data remains an open question.

While fNIRS differs from fMRI in its spatial specificity and instrumentation, both techniques are well-suited to measure the task-evoked hemodynamic response in the cerebral

cortex. As these methods are believed to measure the same underlying biological signal (Buxton, 2010; Huppert et al., 2006), it is worth noting the clear correspondence between our results and those found in functional MRI. Specifically, the load-dependent increase in fronto-parietal metabolic activity for the 2-back task relative to the 1-back task in our study replicates many similar findings in the fMRI literature (Mencarelli et al., 2019; Owen et al., 2005; Rottschy et al., 2012; Schmidt et al., 2009). Further recent fMRI studies have indicated that well practiced tasks may result in reduced PFC activation as a result of increased neural efficiency (Miró-Padilla et al., 2019; Thompson et al., 2016), which is consistent with the negative relationship between mid-frontal activity and performance on the 1-back task demonstrated by the behavioral PLS analysis in the present study.

The current study has a few limitations. Specifically, though it would have been ideal to get clean data from bilateral frontal and parietal cortex, the montage was only able to cover the right parietal cortex and the overall the signal-to-noise ratio (SNR) in these channels was very poor compared to the frontal channels. By design, the Brain AnalyzIR Toolbox performs robust regression to downweight outliers and low SNR channels, which is necessary to avoid high false positive rates (Barker et al., 2013; Santosa et al., 2018). However, it is unclear whether the weaker effects in this area are due to lack of good signal (a false negative) or actually due to a lack of parietal cortical involvement in the N-back task relative to prefrontal cortex. Relatedly, this robust regression and conservative multiple comparisons correction (using q -values with 43 channels, 3 N-back levels, and both HbO and HbR) may explain the limited number of channels demonstrating statistically significant activation for task relative to baseline. Lastly, as our sample was relatively young (average age = 23.6 years), these results may not generalize to older populations. Future research with a larger age range would be beneficial to examine whether this load-dependent pattern of accuracy and fNIRS activation is altered in an older cohort.

Group-level results as an HRF time series, averaged across 5 general ROIs with 8–10 channels each. Color-coded montage in lower right indicates the channels included in each ROI.

Additionally, it is unclear from these data why the PLS analysis found the load-dependent relationship in deoxyhemoglobin (HbR) concentration changes but not in oxyhemoglobin (HbO). In general, the HbO signal is larger than HbR which makes it easier to detect significant effects in task-based fNIRS, as is the case in this study's N-back contrasts. Relative to HbO, the HbR signal is slower and more tightly coupled with the BOLD response of fMRI (Huppert et al., 2006). Therefore, one possibility is that the relationship between task performance and activation got stronger with increasing time performing the task, which may be reflected to a larger extent in the slower HbR signal. Additionally, as HbR is more sensitive to oxygen metabolism (and HbO is more sensitive to blood flow changes), it may be that this effect is more driven by metabolic changes. However, future investigations would be required to directly test these possibilities.

Only significant channels ($q < 0.05$) are shown. Channels are displayed on top of 10–20 coordinates and grayscale depth maps for left and right Inferior Frontal Gyri, medial Superior Frontal Gyri, and right Superior and Inferior Parietal Gyri. For HbO contrasts,

positive t -values (red) correspond to relatively larger activity for the first term in the contrast, and negative t -values (blue) correspond to larger activity for the second term. The opposite pattern applies to HbR contrasts.

In conclusion, the present study demonstrates that fNIRS activation is sensitive to cognitive load and is differentially affected by performance based on task difficulty. This work demonstrates the efficacy of using robust statistical procedures to deal with unique statistical properties of fNIRS signals and the utility of implementing data-driven, multivariate techniques to elucidate more nuanced relationships between brain activity and behavior.

Acknowledgments

Supported by the National Science Foundation BCS-1632445 to M.G.B. the TFK Foundation and the John Templeton Foundation to M.G.B. (The University of Chicago Center for Practical Wisdom and the Virtue, Happiness and Meaning of Life Scholars group) the Grossman Institute for Neuroscience, Quantitative Biology and Human Behavior Shared Equipment Award to M.G.B.); the Mansueto Institute for Urban Innovation [Postdoctoral Fellowship to K.W.C.].

Our fNIRS device was provided by the University of Chicago Grossman Institute for Neuroscience, Quantitative Biology, and Human Behavior (Shared Equipment Award).

We thank Jaime Young, Tanvi Laktahkia, Olivia Paraschos, and Anabella Pinton for their assistance in data collection.

References

- Aghajani H, Garbey M, Omurtag A, 2017. Measuring mental workload with EEG+ fNIRS. *Front. Hum. Neurosci* 11, 359. [PubMed: 28769775]
- Awh E, Jonides J, Smith EE, Schumacher EH, Koeppel RA, Katz S, 1996. Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. *Psychol. Sci* 7 (1), 25–31.
- Ayaz H, Shewokis PA, Bunce S, Izzetoglu K, Willems B, Onaral B, 2012. Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59 (1), 36–47. [PubMed: 21722738]
- Barker JW, Aarabi A, Huppert TJ, 2013. Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomed. Opt. Express* 4 (8), 1366–1379. [PubMed: 24009999]
- Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol* 57 (1), 289–300.
- Berman MG, Mistic B, Buschkuhl M, Kross E, Deldin PJ, Peltier S, Churchill NW, Jaeggi SM, Vakorin V, McIntosh AR, Jonides J, 2014. Does resting-state connectivity reflect depressive rumination? A tale of two analyses. *NeuroImage* 103, 267–279. [PubMed: 25264228]
- Bialystok E, Craik FIM, Grady C, Chau W, Ishii R, Gunji A, Pantev C, 2005. Effect of bilingualism on cognitive control in the Simon task: evidence from MEG. *Neuroimage* 24 (1), 40–49. [PubMed: 15588595]
- Bunce SC, Izzetoglu K, Ayaz H, Shewokis P, Izzetoglu M, Pourrezaei K, Onaral B, 2011. Implementation of fNIRS for monitoring levels of expertise and mental workload. *Foundations of augmented cognition. Directing the future of adaptive systems* 13–22.
- Buracas GT, Boynton GM, 2002. Efficient design of event-related fMRI experiments using M-sequences. *Neuroimage* 16 (3 Pt 1), 801–813. [PubMed: 12169264]
- Buxton RB, 2010. Interpreting oxygenation-based neuroimaging signals: the importance and the challenge of understanding brain oxygen metabolism. *Front. Neuroenergetics* 2, 8. [PubMed: 20616882]

- Buxton RB, 2013. The physics of functional magnetic resonance imaging (fMRI). *Rep. Prog. Phys* 76 (9), 096601. [PubMed: 24006360]
- Causse M, Chua Z, Peysakhovich V, Del Campo N, Matton N, 2017. Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Sci. Rep* 7 (1), 5222. [PubMed: 28701789]
- Chang H, Sprute L, Maloney EA, Beilock SL, Berman MG, 2017. Simple arithmetic: not so simple for highly math anxious individuals. *Soc. Cogn. Affect. Neurosci* 12 (12), 1940–1949. [PubMed: 29140499]
- Choe KW, Blake R, Lee S–H, 2014a. Dissociation between neural signatures of stimulus and choice in population activity of Human V1 during perceptual decision-making. *J. Neurosci* 34 (Issue 7), 2725–2743. doi: 10.1523/jneurosci.1606-13.2014. [PubMed: 24523561]
- Choe KW, Blake R, Lee S–H, 2014b. Dissociation between neural signatures of stimulus and choice in population activity of human V1 during perceptual decision-making. *J. Neurosci* 34 (7), 2725–2743. [PubMed: 24523561]
- Choe KW, Blake R, Lee S–H, 2016. Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vis. Res* 118, 48–59. [PubMed: 25578924]
- Churchill NW, Spring R, Grady C, Cimprich B, Askren MK, Reuter-Lorenz PA, Jung MS, Peltier S, Strother SC, Berman MG, 2016. The suppression of scale-free fMRI brain dynamics across three different sources of effort: aging, task novelty and task difficulty. *Sci. Rep* 6, 30895. [PubMed: 27498696]
- Conway ARA, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW, 2005. Working memory span tasks: a methodological review and user’s guide. *Psychon. Bull. Rev* 12 (5), 769–786. [PubMed: 16523997]
- Cooper RJ, Selb J, Gagnon L, Phillip D, Schytz HW, Iversen HK, Ashina M, Boas DA, 2012. A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Front. Neurosci* 6, 147. [PubMed: 23087603]
- Cui X, Bray S, Bryant DM, Glover GH, Reiss AL, 2011. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage* 54 (4), 2808–2821. [PubMed: 21047559]
- Dunst B, Benedek M, Jauk E, Bergner S, Koschutnig K, Sommer M, Ischebeck A, Spinath B, Arendasy M, Bühner M, Freudenthaler H, Neubauer AC, 2014. Neural efficiency as a function of task demands. *Intelligence* 42 (100), 22–30. [PubMed: 24489416]
- Fishburn FA, Norr ME, Medvedev AV, Vaidya CJ, 2014. Sensitivity of fNIRS to cognitive state and load. *Front. Hum. Neurosci* 8, 76. [PubMed: 24600374]
- Herff C, Heger D, Fortmann O, Hennrich J, Putze F, Schultz T, 2014. Mental workload during n-back task—Quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci* 7. doi: 10.3389/fnhum.2013.00935. [PubMed: 24478673]
- Hocke LM, Oni IK, Duszynski CC, Corrigan AV, Frederick BD, Dunn JF, 2018. Automated Processing of fNIRS Data—a visual guide to the pitfalls and consequences. *Algorithms* 11 (5). doi: 10.3390/a11050067.
- Hosseini SMH, Bruno JL, Baker JM, Gundran A, Harbott LK, Gerdes JC, Reiss AL, 2017. Neural, physiological, and behavioral correlates of visuomotor cognitive load. *Sci. Rep* 7 (1), 8866. [PubMed: 28821719]
- Huppert TJ, 2016. Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics* 3 (1), 010401. [PubMed: 26989756]
- Huppert TJ, Diamond SG, Franceschini MA, Boas DA, 2009. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt* 48 (10), D280–D298. [PubMed: 19340120]
- Huppert TJ, Hoge RD, Diamond SG, Franceschini MA, Boas DA, 2006. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29 (2), 368–382. [PubMed: 16303317]
- Jack AI, Shulman GL, Snyder AZ, McAvoy M, Corbetta M, 2006. Separate modulations of human V1 associated with spatial attention and task structure. *Neuron* 51 (1), 135–147. [PubMed: 16815338]

- Kane MJ, Conway ARA, Miura TK, Colflesh GJH, 2007. Working memory, attention control, and the N-back task: a question of construct validity. *J. Exp. Psychol. Learn. Mem. Cogn* 33 (3), 615–622. [PubMed: 17470009]
- Kardan O, Adam KCS, Mance I, Churchill NW, Vogel EK, Berman MG, 2020a. Distinguishing cognitive effort and working memory load using scale-invariance and alpha suppression in EEG. *Neuroimage* 211, 116622. [PubMed: 32068164]
- Kardan O, Layden E, Choe KW, Lyu M, Zhang X, Beilock SL, Rosenberg MD, & Berman MG, 2020. Scale-invariance in brain activity predicts practice effects in cognitive performance. In bioRxiv (p. 2020.05.25.114959). doi: 10.1101/2020.05.25.114959.
- Kirchner WK, 1958. Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol* 55 (4), 352–358. [PubMed: 13539317]
- Klem GH, Lüders HO, Jasper HH, Elger C, 1999. The ten-twenty electrode system of the International Federation. *The International Federation of Clinical Neurophysiology. Electroencephalogr. Clin. Neurophysiol. Suppl* 52, 3–6. [PubMed: 10590970]
- Krishnan A, Williams LJ, McIntosh AR, Abdi H, 2011. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* 56 (2), 455–475. [PubMed: 20656037]
- Kuruvilla MS, Green JR, Ayaz H, Murman DL, 2013. Neural correlates of cognitive decline in ALS: an fNIRS study of the prefrontal cortex. *Cogn. Neurosci* 4 (2), 115–121. [PubMed: 24073736]
- Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT, 2000. Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp* 10 (3), 120–131. [PubMed: 10912591]
- Lawrence MA, 2016. ez: Easy Analysis and Visualization of Factorial Experiments. <https://CRAN.R-project.org/package=ez>.
- Lobaugh NJ, West R, McIntosh AR, 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology* 38 (3), 517–530. [PubMed: 11352141]
- Lüdecke D, 2020. sjstats: Statistical Functions for Regression Models (Version 0.17.9). doi: 10.5281/zenodo.1284472.
- Mandrick K, Derosiere G, Dray G, Coulon D, Micallef J–P, Perrey S, 2013. Prefrontal cortex activity during motor tasks with additional mental load requiring attentional demand: a near-infrared spectroscopy study. *Neurosci. Res* 76 (3), 156–162. [PubMed: 23665138]
- Mandrick K, Peysakhovich V, Rémy F, Lepron E, Causse M, 2016. Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biol. Psychol* 121 (Pt A), 62–73. [PubMed: 27725244]
- Matsuda G, Hiraki K, 2006. Sustained decrease in oxygenated hemoglobin during video games in the dorsal prefrontal cortex: a NIRS study of children. *NeuroImage* 29 (3), 706–711. [PubMed: 16230030]
- McIntosh AR, Kovacevic N, Itier RJ, 2008. Increased brain signal variability accompanies lower behavioral variability in development. *PLoS Comput. Biol* 4 (7), e1000106. [PubMed: 18604265]
- McIntosh AR, Lobaugh NJ, 2004. Partial least squares analysis of neuroimaging data: applications and advances. *Neuroimage* 23 Suppl 1, S250–S263. [PubMed: 15501095]
- McKendrick R, Ayaz H, Olmstead R, Parasuraman R, 2014. Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with NIRS. *NeuroImage* 85 (Pt 3), 1014–1026. [PubMed: 23727530]
- Mencarelli L, Neri F, Momi D, Menardi A, Rossi S, Rossi A, Santarnecchi E, 2019. Stimuli, presentation modality, and load-specific brain activity patterns during n-back task. *Hum. Brain Mapp* 40 (13), 3810–3831. [PubMed: 31179585]
- Meule A, 2017. Reporting and interpreting working memory performance in n-back tasks. *Front. Psychol* 8, 352. [PubMed: 28326058]
- Miró-Padilla A, Bueichekú E, Ventura-Campos N, Flores-Compañ M–J, Parcet MA, Ávila C, 2019. Long-term brain effects of N-back training: an fMRI study. *Brain Imaging Behav.* 13 (4), 1115–1127. [PubMed: 30006860]
- Moore ME, Ross BM, 1963. Context effects in running memory. *Psychol. Rep* 12 (2), 451–465.

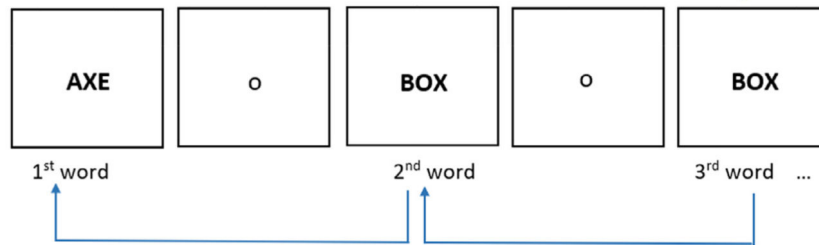
- Morais GAZ, Balardin JB, Sato JR, 2018. fNIRS Optodes' Location Decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. *Sci. Rep* 8 (1), 3341. [PubMed: 29463928]
- Murata A, Park J, Kovelman I, Hu X, Kitayama S, 2015. Culturally non-preferred cognitive tasks require compensatory attention: a functional near infrared spectroscopy (fNIRS) investigation. *Cult. Brain* 3 (1), 53–67.
- Neubauer AC, Fink A, 2009. Intelligence and neural efficiency. *Neurosci. Biobehav. Rev* 33 (7), 1004–1023. [PubMed: 19580915]
- Oberauer K, 2005. Binding and inhibition in working memory: individual and age differences in short-term recognition. *J. Exp. Psychol. Gen* 134 (3), 368–387. [PubMed: 16131269]
- Owen AM, McMillan KM, Laird AR, Bullmore E, 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp* 25 (1), 46–59. [PubMed: 15846822]
- Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK, 2019. PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51 (1), 195–203. [PubMed: 30734206]
- Pinti P, Scholkmann F, Hamilton A, Burgess P, Tachtsidis I, 2018a. Current status and issues regarding pre-processing of fNIRS neuroimaging data: an investigation of diverse signal filtering methods within a general linear model framework. *Front. Hum. Neurosci* 12, 505. [PubMed: 30687038]
- Pinti P, Tachtsidis I, Hamilton A, Hirsch J, Aichelburg C, Gilbert S, Burgess PW, 2018b. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann. N. Y. Acad. Sci* 10.1111/nyas.13948.
- Ravizza SM, Delgado MR, Chein JM, Becker JT, Fiez JA, 2004. Functional dissociations within the inferior parietal cortex in verbal working memory. *Neuroimage* 22 (2), 562–573. [PubMed: 15193584]
- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rolls ET, Joliot M, Tzourio-Mazoyer N, 2015. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage* 122, 1–5. [PubMed: 26241684]
- Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, Fox PT, Eickhoff SB, 2012. Modelling neural correlates of working memory: a coordinate-based meta-analysis. *Neuroimage* 60 (1), 830–846. [PubMed: 22178808]
- Santosa H, Fishburn F, Zhai X, Huppert TJ, 2019. Investigation of the sensitivity-specificity of canonical- and deconvolution-based linear models in evoked functional near-infrared spectroscopy. *Neurophotonics* 6 (2), 025009. [PubMed: 31172019]
- Santosa H, Zhai X, Fishburn F, Huppert T, 2018. The NIRS Brain AnalyzIR Toolbox. *Algorithms* 11 (5), 73.
- Sato H, Yahata N, Funane T, Takizawa R, Katura T, Atsumori H, Nishimura Y, Kinoshita A, Kiguchi M, Koizumi H, Fukuda M, Kasai K, 2013. A NIRS-fMRI investigation of prefrontal cortex activity during a working memory task. *Neuroimage* 83, 158–173. [PubMed: 23792984]
- Sayala S, Sala JB, Courtney SM, 2006. Increased neural efficiency with repeated performance of a working memory task is information-type dependent. *Cereb. Cortex* 16 (5), 609–617. [PubMed: 16079245]
- Schmidt H, Jogia J, Fast K, Christodoulou T, Haldane M, Kumari V, Frangou S, 2009. No gender differences in brain activation during the N-back task: an fMRI study in healthy individuals. *Hum. Brain Mapp.* 30 (11), 3609–3615. [PubMed: 19387979]
- Scholkmann F, Kleiser S, Metz AJ, Zimmermann R, Mata Pavia J, Wolf U, Wolf M, 2014. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* (85 Pt 1) 6–27. [PubMed: 23684868]
- Sirotin YB, Cardoso M, Lima B, Das A, 2012. Spatial homogeneity and task-synchrony of the trial-related hemodynamic signal. *Neuroimage* 59 (3), 2783–2797. [PubMed: 22036678]
- Strangman G, Franceschini MA, Boas DA, 2003. Factors affecting the accuracy of near-infrared spectroscopy concentration calculations for focal changes in oxygenation parameters. *Neuroimage* 18 (4), 865–879. [PubMed: 12725763]

- Thompson TW, Waskom ML, Gabrieli JDE, 2016. Intensive working memory training produces functional changes in large-scale frontoparietal networks. *J. Cogn. Neurosci* 28 (4), 575–588. [PubMed: 26741799]
- Tong Y, Lindsey KP, deB Frederick B, 2011. Partitioning of physiological noise signals in the brain with concurrent near-infrared spectroscopy and fMRI. *J. Cerebr. Blood Flow Metabol.: Offic. J. Int. Soc. Cerebr. Blood Flow Metabol* 31 (12), 2352–2362.
- Turner BO, Paul EJ, Miller MB, Barbey AK, 2018. Small sample sizes reduce the replicability of task-based fMRI studies. *Commun. Biol* 1, 62. [PubMed: 30271944]
- Yücel MA, Selb JJ, Huppert TJ, Franceschini MA, Boas DA, 2017. Functional Near infrared spectroscopy: enabling routine functional brain imaging. *Curr. Opin. Biomed. Eng* doi: 10.1016/j.cobme.2017.09.011.

1-BACK EXAMPLE

Press "n" because this word (BOX) does *not match* the word presented 1 trial back (AXE)

Press "m" because this word (BOX) does *match* the word presented 1 trial back (BOX)



3-BACK EXAMPLE

Here, you would press "m" because this word (AXE) does *match* the word presented 3 trials back (AXE)

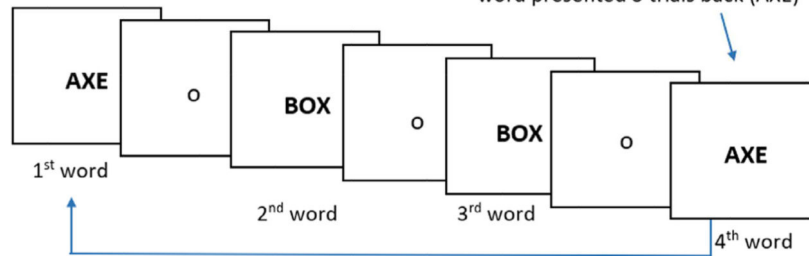


Fig. 1. N-back Task. Example of 1-back task (Top) and 3-back (Bottom). 2-back task not shown.

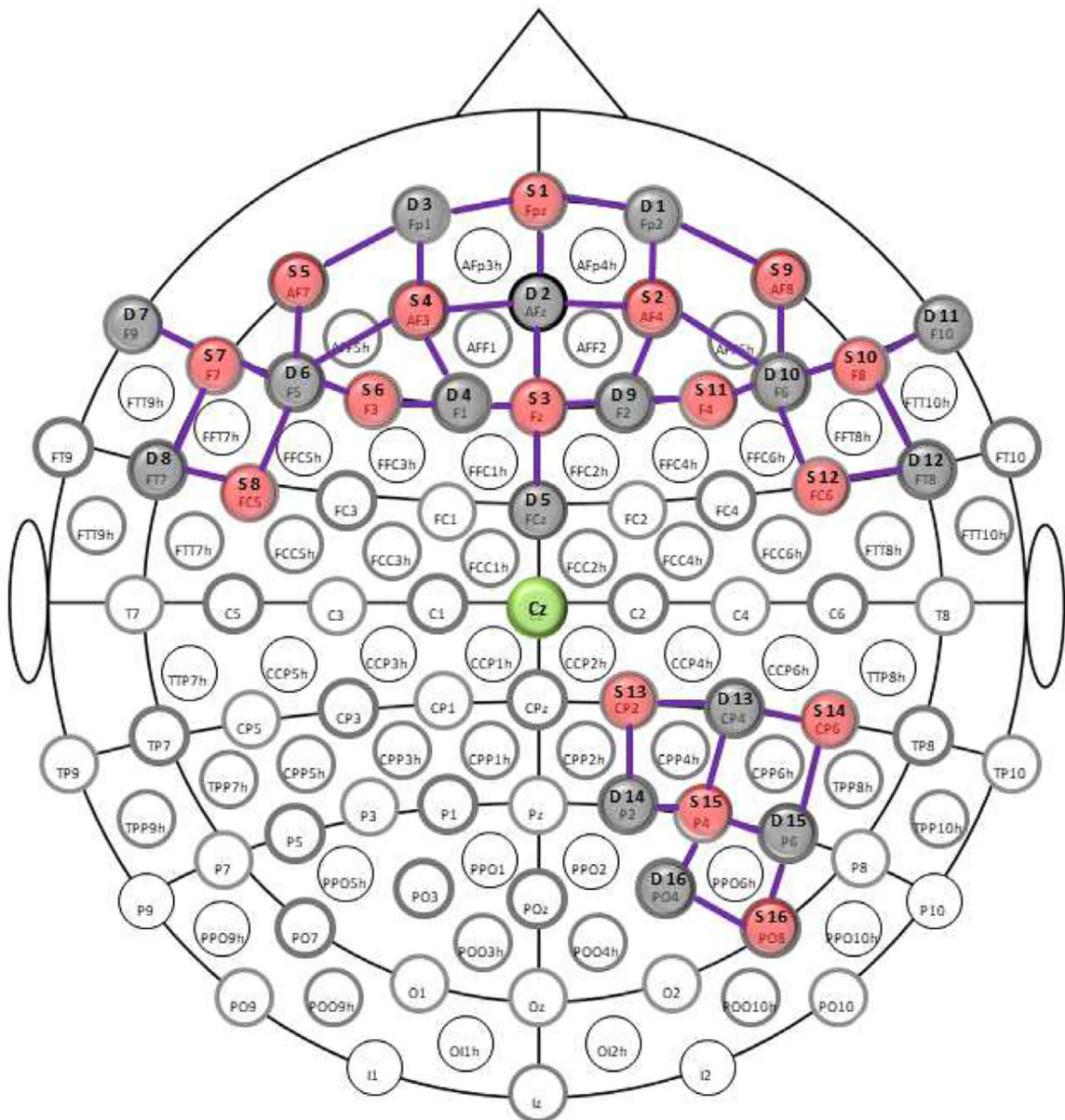


Fig. 2. fNIRS Montage in international 10–10 coordinate space. Montage with 16 × 16 frontal source-detector pairs and 4 × 4 right parietal source-detector pairs. Sources are indicated in red, detectors are indicated in gray, and channels are indicated by purple lines. Cz highlighted in green.

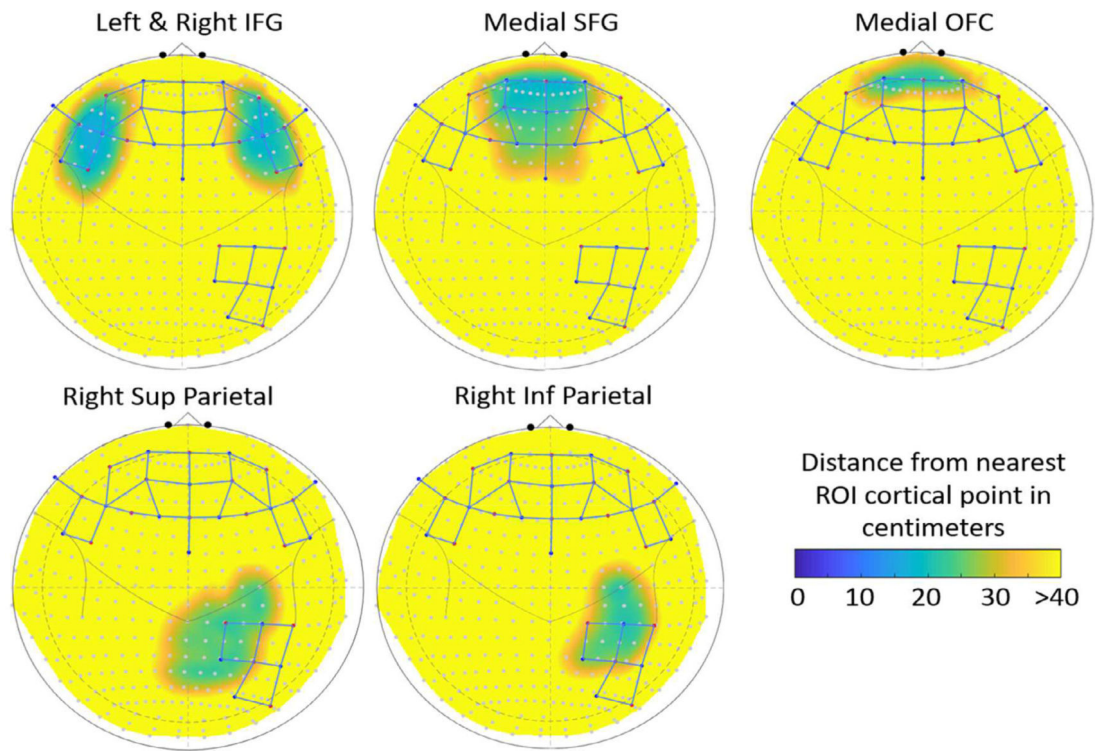


Fig. 3. Gross ROI depth maps with superimposed montage. fNIRS montage (registered to Colin27 atlas) and depth map for 6 ROIs taken from the talairach daemon parcellation: Left and Right Inferior Frontal Gyrus, Medial Superior Frontal Gyrus, Medial Orbitofrontal Cortex, Right Superior Parietal Gyrus, Right Inferior Parietal Gyrus.

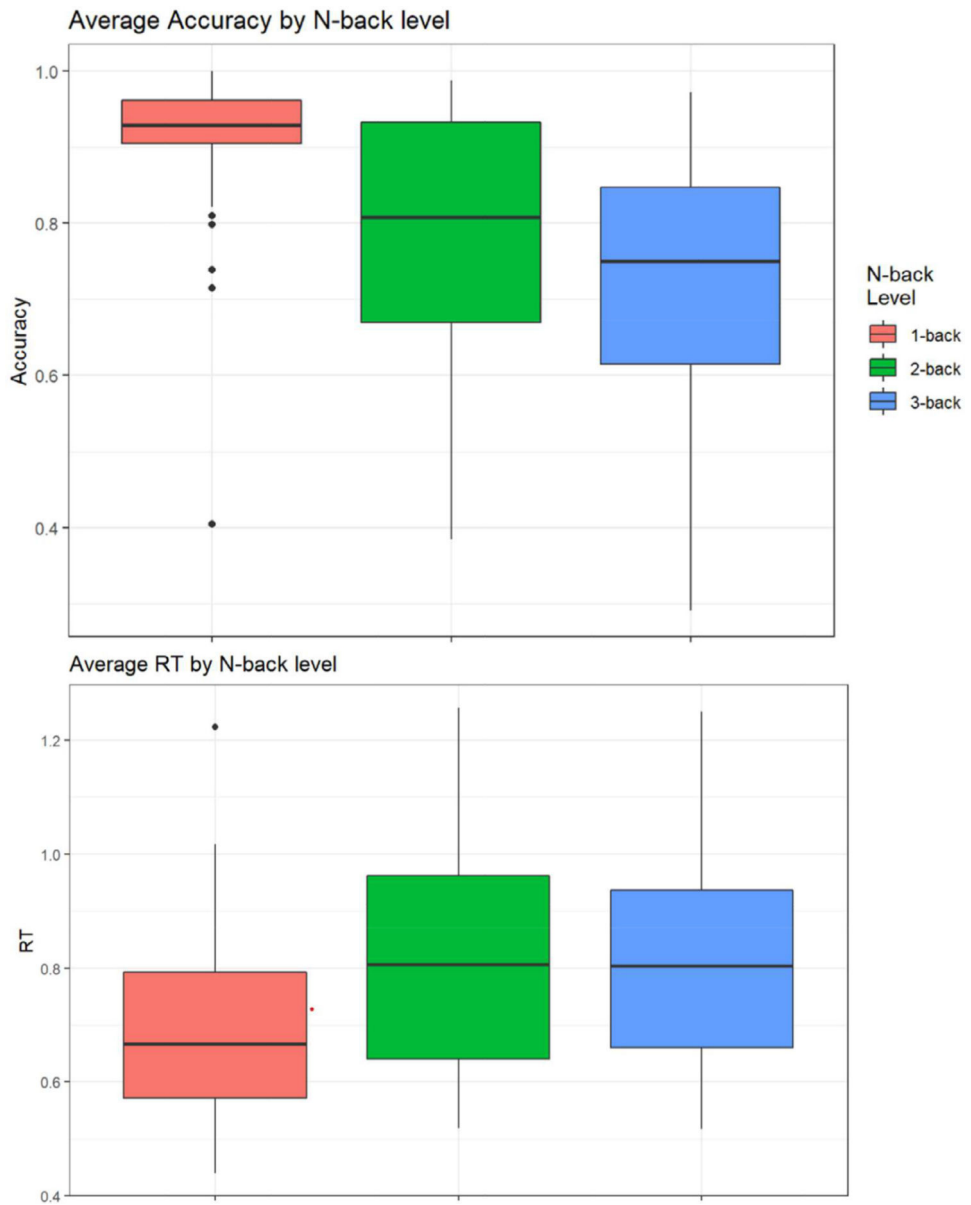


Fig. 4. Boxplots of Average Accuracy & RT by N-back Level for all participants.

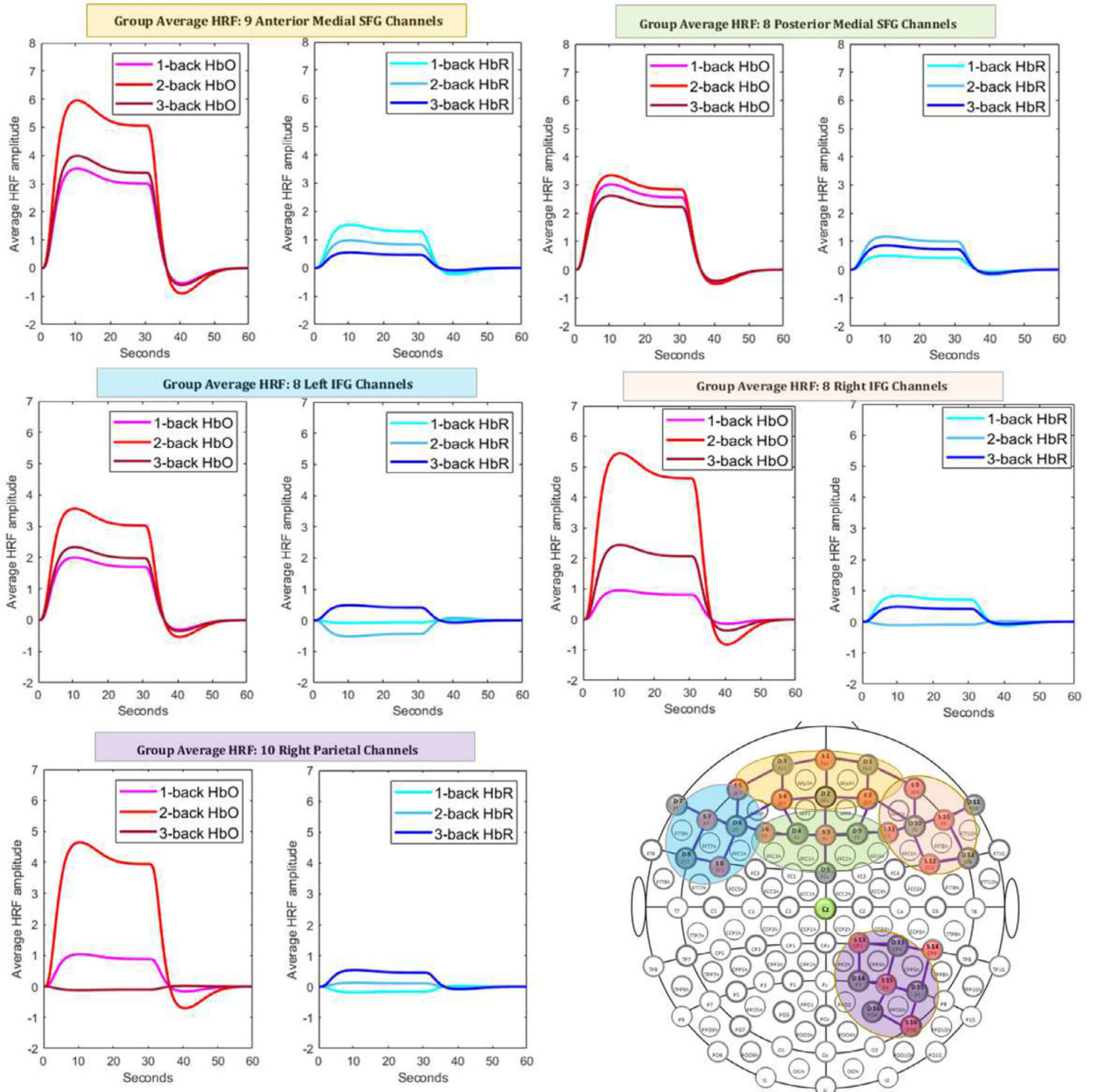


Fig. 5. Group Average HRF plots by N-back level. Group-level results as an HRF time series, averaged across 5 general ROIs with 8–10 channels each. Color-coded montage in lower right indicates the channels included in each ROI.

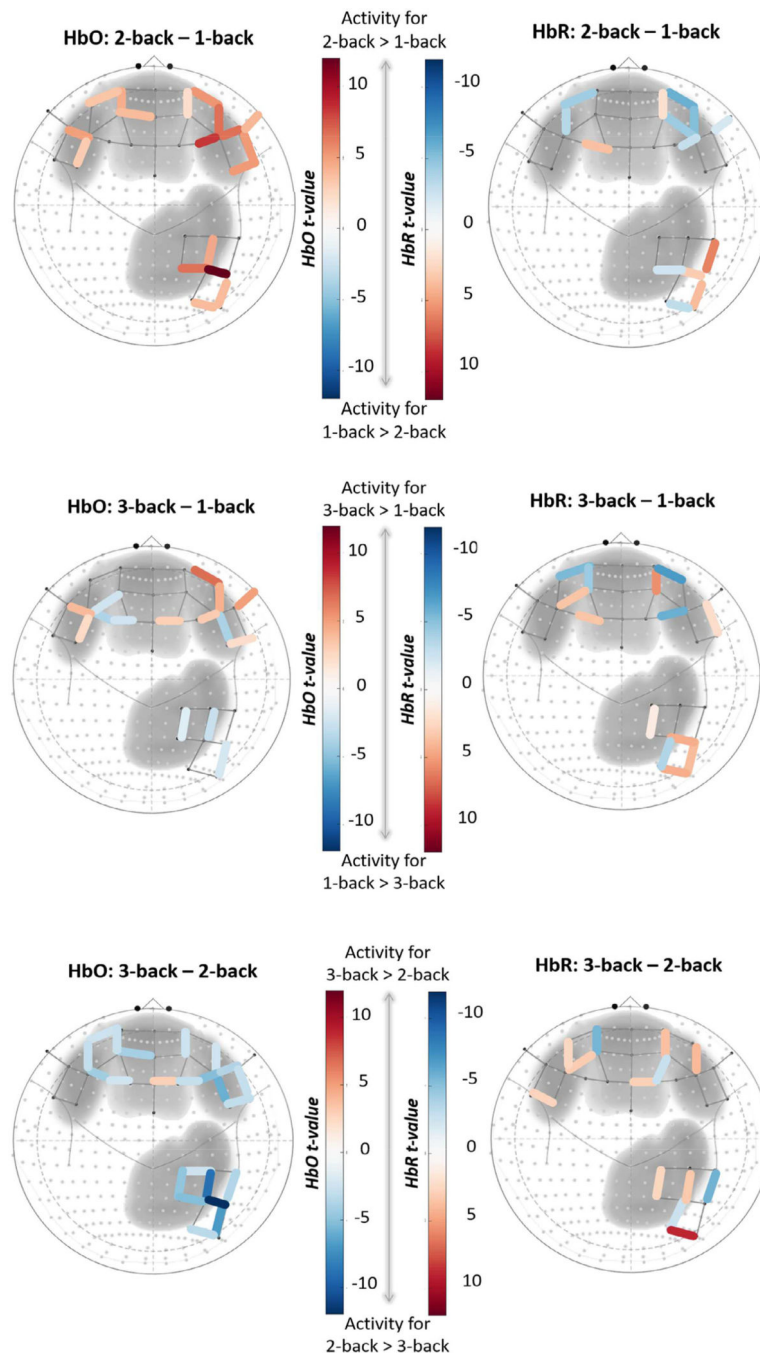


Fig. 6. N-back level contrasts for HbO (left) and HbR (right). Only significant channels ($q < 0.05$) are shown. Channels are displayed on top of 10–20 coordinates and grayscale depth maps for left and right Inferior Frontal Gyri, medial Superior Frontal Gyri, and right Superior and Inferior Parietal Gyri. For HbO contrasts, positive t -values (red) correspond to relatively larger activity for the first term in the contrast, and negative t -values (blue) correspond to larger activity for the second term. The opposite pattern applies to HbR contrasts.

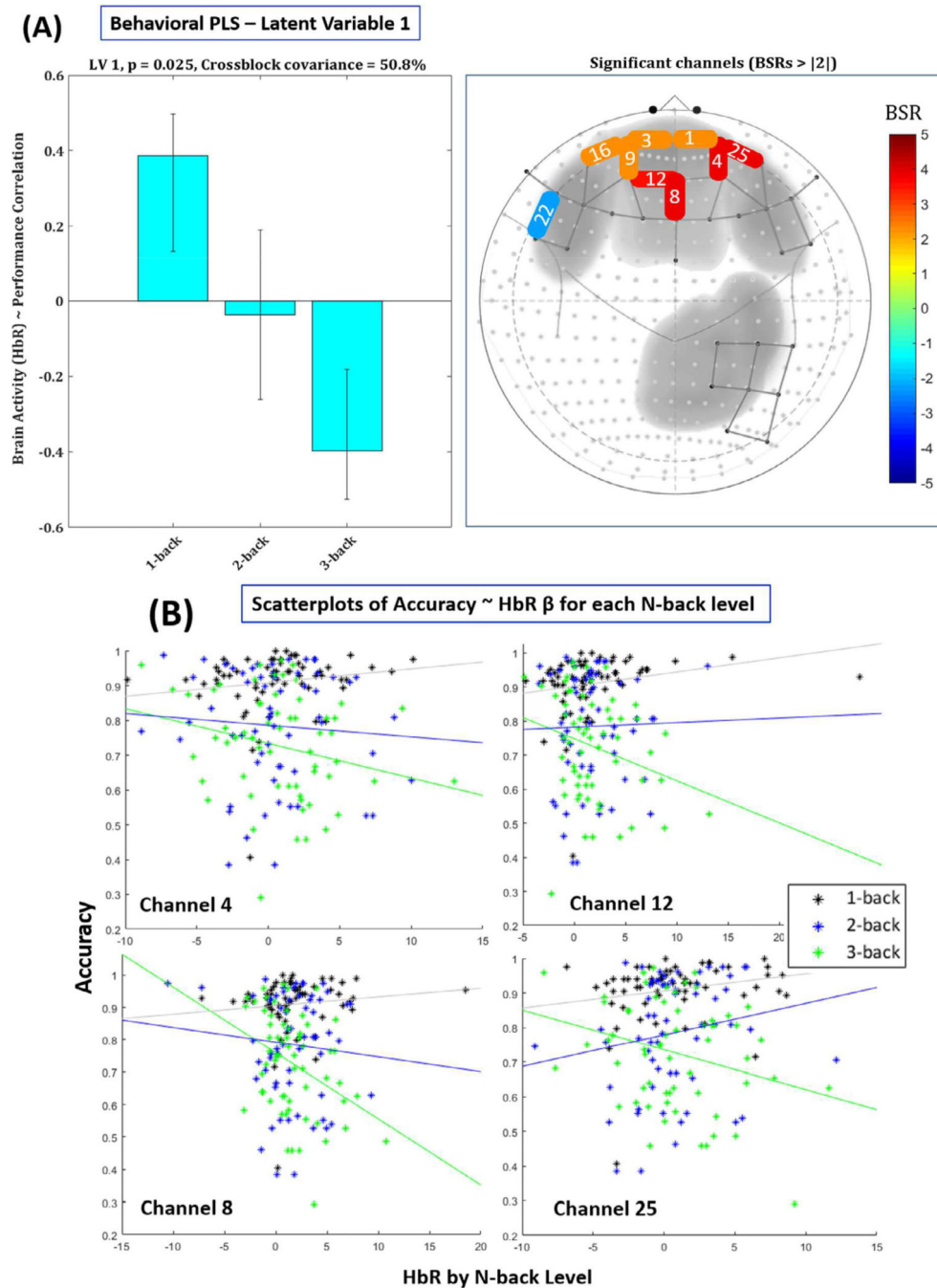


Fig. 7. LV 1 demonstrated an N-back load-dependent relationship between changes in deoxyhemoglobin concentrations (HbR) and performance. (A) The left panel shows correlation between accuracy and HbR concentration change separately by N-back level. Error bars are 95% confidence intervals around the mean correlation. The right panel shows channels (labeled by number), which had bootstrap ratios (BSR) > |2|. (B) Scatterplots

showing the correlation between HbR (β for task-evoked change from baseline) and performance (accuracy) at 4 channels with BSRs > 3, separated by N-back level.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Significant Activation by Channel & ROI for each N-back level.

Significant channels (Source (S) - Detector (D) pairs) identified as FDR-corrected $q < 0.05$. ROI defined by maximal coverage of talairach daemon parcellation ROI. p -value listed is before FDR correction. Power listed is the estimated type-II power for that entry (calculated by computing the minimum detectable change as detailed in Harcum & Dressing, 2015).

	S	D	ROI	<i>t</i> -stat	<i>p</i>	<i>q</i>	power
1-back HbO	1	2	L Medial Superior Frontal Gyrus	3.24	0.001	0.041	0.77
	2	1	R Medial Superior Frontal Gyrus	3.48	0.001	0.032	0.83
2-back HbO	4	2	L Medial Superior Frontal Gyrus	3.28	0.001	0.042	0.78
	4	3	L Medial Superior Frontal Gyrus	3.66	< 0.001	0.023	0.88
	2	10	R Inferior Frontal Gyrus	3.3	0.001	0.042	0.79
3-back HbO	11	10	R Inferior Frontal Gyrus	3.86	< 0.001	0.017	0.91
	15	15	R Inferior Parietal	4.41	< 0.001	0.004	0.97
	-	-	-	-	-	-	-

Table 2
Significant Channels for LV 1

Channel number based on source (S) - detector (D) pair. ROI label defined by maximal coverage of talairach daemon ROI. Channels ordered by size of bootstrap ratio. Bootstrap ratios $> |2|$ were considered significant.

Channel #	S	D	ROI	Bootstrap Ratio
25	9	1	R Middle/Superior Frontal Gyrus	4.3
8	3	2	Medial Superior Frontal Gyrus	3.8
12	4	2	L Superior Frontal Gyrus	3.5
4	2	1	R Superior Frontal Gyrus	3.3
16	5	3	L Middle/Superior Frontal Gyrus	2.6
1	1	1	Medial Superior Frontal Gyrus	2.3
3	1	3	Medial Superior Frontal Gyrus	2.3
9	3	4	Medial Superior Frontal Gyrus	2.2
22	7	8	L Inferior Frontal Gyrus	-2.4

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript