



Research Report

Scale invariance in fNIRS as a measurement of cognitive load



Chu Zhuang^{a,1}, Kimberly L. Meidenbauer^{a,*}, Omid Kardan^a,
 Andrew J. Stier^a, Kyoung Whan Choe^{a,b}, Carlos Cardenas-Iniguez^a,
 Theodore J. Huppert^c and Marc G. Berman^{a,d,**}

^a Environmental Neuroscience Lab, Department of Psychology, The University of Chicago, USA

^b Mansueto Institute for Urban Innovation, The University of Chicago, USA

^c Department of Electrical and Computer Engineering, The University of Pittsburgh, USA

^d Neuroscience Institute, The University of Chicago, USA

ARTICLE INFO

Article history:

Received 24 September 2021

Reviewed 10 January 2022

Revised 29 April 2022

Accepted 23 May 2022

Action editor Kate Hoy

Published online 31 May 2022

Keywords:

Functional near-infrared
 spectroscopy

Hurst exponent

N-back task

Partial least squares

Task difficulty

ABSTRACT

Scale invariant neural dynamics are a relatively new but effective means of measuring changes in brain states as a result of varied cognitive load and task difficulty. This study tests whether scale invariance (as measured by the Hurst exponent, H) can be used with functional near-infrared spectroscopy (fNIRS) to quantify cognitive load, paving the way for scale-invariance to be measured in a variety of real-world settings. We analyzed H extracted from the fNIRS time series while participants completed an N-back working memory task. Consistent with what has been demonstrated in fMRI, the current results showed that scale-invariance analysis significantly differentiated between task and rest periods as calculated from both oxy- (HbO) and deoxy-hemoglobin (HbR) concentration changes. Results from both channel-averaged H and a multivariate partial least squares approach (Task PLS) demonstrated higher H during the 1-back task than the 2-back task. These results were stronger for H derived from HbR than from HbO. This suggests that scale-free brain states are a robust signature of cognitive load and not limited by the specific neuroimaging modality employed. Further, as fNIRS is relatively portable and robust to motion-related artifacts, these preliminary results shed light on the promising future of measuring cognitive load in real life settings.

© 2022 Elsevier Ltd. All rights reserved.

* Corresponding author. Environmental Neuroscience Lab, Department of Psychology, The University of Chicago, 5848 S University Avenue, Chicago, IL 60637, USA.

** Corresponding author. Environmental Neuroscience Lab, Department of Psychology, The University of Chicago, 5848 S University Avenue, Chicago, IL 60637, USA.

E-mail addresses: meidenbauer@uchicago.edu (K.L. Meidenbauer), bermanm@uchicago.edu (M.G. Berman).

¹ Authors contributed equally.

<https://doi.org/10.1016/j.cortex.2022.05.009>

0010-9452/© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

Functional near-infrared spectroscopy (fNIRS) is a neuroimaging technique that has gained increasing attention in recent years due to its relative robustness to motion artifacts and environmental noise, making it more suitable for neuroimaging outside standard laboratory settings (Huppert et al., 2006; Pinti et al., 2018) and with hard-to-test populations (Lloyd-Fox et al., 2010). Like functional MRI, fNIRS measures changes in the brain's hemodynamic response, but does so using light spectroscopy at near-infrared wavelengths. At the onset of regional neural activity, metabolic demands rise, and blood flow increases in that area. This increased blood flow leads to higher concentrations of oxygenated hemoglobin concentrations (HbO) and lower concentrations of deoxy-hemoglobin concentrations (HbR). Measuring these oxy- and deoxy-hemoglobin concentration changes in fNIRS provide a complementary method to fMRI with lower cost and greater use-case flexibility (Ferrari & Quaresima, 2012). Many findings from fMRI have been replicated using fNIRS, such as activation changes that result from varying cognitive load in working memory tasks (Cui et al., 2011; Mencarelli et al., 2019; Owen et al., 2005).

Cognitive load refers to the level of demand and difficulty people bear when performing a cognitive task (Garbarino & Edell, 1997; Paas & Van Merriënboer, 1993). More specifically, it often indicates the amount of working memory used or the number of items people are actively holding in working memory (Mencarelli et al., 2019; Owen et al., 2005). However, when people are continuously under high cognitive load and there are sustained demands on attention and working memory, performance can decrease, and people may suffer from high levels of mental fatigue. This fatigue can be very dangerous, especially for those who work in positions requiring sustained attention, such as pilots, doctors, technicians under military duties etc (Klaassen et al., 2014; Pattyn et al., 2008). Similarly, in an educational setting, children may not learn effectively or perform well if they are already experiencing high levels of cognitive fatigue (Sievertsen et al., 2016).

Both behavioral (i.e., speed and accuracy in a task) and physiological measures (i.e., skin conductance, blood glucose, cardiac function) have been used to measure cognitive load in previous work (Kennedy & Scholey, 2000; Naccache et al., 2005). However, behavioral measures do not directly reflect cognitive load and may neglect 'compensatory effort' (Kardan et al., 2020a). That is, people may exert extra effort and experience higher cognitive load and fatigue while maintaining the same level of performance. Physiological methods do not index cognitive load directly as well, as they measure cardiovascular and sympathetic nervous activities, which are an indirect proxy of cognitive load. It can be hard to tell what changes may be predominantly due to cognitive effort and what might be primarily due to arousal (Fishburn et al., 2014) and stress (Grady et al., 2010; McKiernan et al., 2003). Thus, behavioral and physiological methods are neither specific nor fully sensitive to detect changes in cognitive load. To better monitor these changes, it would be advantageous to identify a reliable neural measure, both sensitive and specific to

cognitive difficulty, effort, or fatigue, which utilizes mobile neuroimaging technology.

Previous neuroimaging studies investigating cognitive load with fNIRS and fMRI have often centered around localized activation that differs based on task difficulty or working memory load (Fishburn et al., 2014; Grady et al., 2010; McKiernan et al., 2003). However, the present study aimed to measure levels of cognitive load through a whole brain neural signature, scale invariance of the broadband brain signal, quantified by the Hurst exponent (H). This signature has been validated with fMRI and EEG studies and has proved sensitive and specific to changes of cognitive load (Churchill et al., 2016; Kardan et al., 2020a; Maxim et al., 2005). H is considered a global measure as changes in scaling due to task demands or individual differences have been found across the whole brain (in fMRI which provides whole brain coverage) and H changes are unidirectional, following the same pattern of effects across regions examined (Churchill et al., 2015; Kardan et al., 2020b). Changes in H are stronger in some regions than others, particularly those that are task-relevant and also show changes in activation and/or functional connectivity due to task demands (Churchill et al., 2016; Kardan et al., 2020b). Thus, H is both a unidirectional signature and shows a certain level of spatial selectivity.

Scale invariance (also called scale-free or fractal scaling) refers to the property of a time series signal that it has persistent autocorrelation with long-range dependency. By this, we mean that similar fluctuation patterns appear at all time scales (whether examined over a short window or a longer window) and no specific frequency band (i.e., alpha band vs beta band, etc.) plays a dominant role. From the frequency perspective, scale invariant neural signals exhibit a power-law relationship between Power Spectral Density of the signal (PSD), and its frequency (f), where $\text{PSD}(f) \sim f^{-\beta}$, $\beta \geq 0$. When the signal is scale-invariant, the slope of this function, β , should be positive (PSD close to $1/f$). The more scale-invariant the signal is, the more inclined the slope is and the higher value of β . Conversely, if the signal is less scale-free, the slope is shallower, with a lower value of β (Kardan et al., 2020b). According to this formula, the degree of temporal scale invariance can be calculated by the Hurst exponent (H), which is related to the slope β via equation $H = (\beta + 1)/2$, or $\beta = 2H - 1$. Therefore, if a signal is scale invariant and shows larger long-range dependency, it will lead to a higher value of H , which corresponds to a higher value of β (steeper slope).

The range of possible values of H are somewhat different however, depending on whether the signal being recorded is stationary (that is, the mean and variance does not change over time as is typical in fMRI; Barnes et al., 2009; Churchill et al., 2016; He, 2011) or nonstationary (as is typically found in neural field potentials or EEG; Freeman & Zhai, 2009; He et al., 2010; Miller et al., 2009). If the signal is stationary, the slope (β) of power spectral density should be between 0 and 1 and the Hurst exponent should fall between .5 and 1 because $H = (\beta + 1)/2$. The floor of H is .5, as this represents white noise, where power is not related to frequency, but rather power is uniformly distributed across frequency. If examined in log space, this would represent a line with zero slope.

In contrast, if the brain signals are nonstationary, then the slope β can exceed 1. For example, in EEG recordings, typical values for β are in the range of 1–2.5, leading to a calculation

of H in the range of approximately 1–1.75 (Dehghani et al., 2010; Freeman et al., 2003). Similarly, nonstationary field potentials yield estimates of H in the range of 2–4 (Eke et al., 2002; Mandelbrot & Van Ness, 1968). Though fNIRS is measuring the same metabolic process as fMRI, fNIRS data does exhibit a certain amount of nonstationarity due to how the data is measured. Specifically, the underlying biological signal in fNIRS is slower than the rate at which the data are sampled can lead to some nonstationarity in the data.

Previous research has found that H , the scale invariance index, is a robust neural indicator of cognitive load (Barnes et al., 2009; Churchill et al., 2016; He, 2011). The effectiveness of H in quantifying cognitive load and task engagement has been validated in both fMRI (BOLD signals) and EEG studies (oscillatory activity; Kardan et al., 2020a). H decreases globally with changes in cognitive load and task difficulty, where it is highest during rest, lower while participants perform tasks, and shows the most suppression during highly cognitively demanding tasks. A defining characteristic of temporal scale-invariance is that this persistent autocorrelation and fractal-like self-similarity is unaffected by the magnitude of the time window examined. As H has been extracted from multiple neuroimaging modalities spanning very different sampling rates and time scales (i.e., .01–.2 Hz in minutes-long fMRI time series, ~30 Hz in seconds-long downsampled EEG time series data), and is not dominated by any particular frequency band (Churchill et al., 2016; Kardan et al., 2020a), this suggests that H is a truly scale-free measure. In other words, slower signals and faster/more frequent signals exhibit the same relationships of higher H indicating lower cognitive effort, and that the Hurst measure applies to all frequencies and not to narrow frequency bands. In this way, this measure is a useful metric to unite different neuroimaging modalities.

It has been proposed that H measures how close the brain is to a “critical state”. Criticality is a concept taken from physics that reflects how readily a system can transition between different states and has been applied to biological systems, such as the brain (Cocchi et al., 2017). In this case, a brain state close to a critical point is minimally stable and maximally sensitive to perturbations from exogenous inputs (Kardan et al., 2020b). In neural network simulations, a critical brain state increases the dynamic range of the system (Gautam et al., 2015; Kinouchi & Copelli, 2006) and facilitates the transfer of information (Boedeker et al., 2012). This framework supports the findings that more scale-free brain states, which are closer to the point of criticality, are associated with exerting less cognitive effort, engaging in an easier task, being younger, or experiencing minimal or no psychopathology symptoms (Churchill et al., 2016; Kardan et al., 2020a, 2020b; Stier et al., 2021). In other words, higher H values reflect brain activity closer to a critical state, allowing higher H brains to take in external information and shift to a different state if needed. H decreases as cognitive load increases as the brain is less able to take in external information needed to transition to a different task-relevant state if required.

However, while the effectiveness of the Hurst exponent as a cognitive load measurement has been validated in both fMRI and EEG studies, it has not yet been examined in fNIRS. Compared with fMRI and EEG, fNIRS is more robust to motion artifacts and environmental noise and is more portable,

making it ideal for investigations in ecologically valid settings (for example, measuring fatigue during driving; Liu et al., 2016, or learning in school aged children; Soltanlou et al., 2018). The results of this study will shed light on future theoretical and practical investigation of cognitive load, fatigue with fNIRS, and even monitoring cognitive load levels and signaling changes by neuro-feedback to mitigate fatigue (Luctkar-Flude and Groll, 2015a). This would allow researchers to measure cognitive fatigue in some difficult scenarios, such as in school settings where children are trying to learn different cognitive tasks, workplace settings such as aeronautics, transportation and even possibly space travel. We expected that, if scale invariant brain states are indeed robust indicators of cognitive load or task difficulty, the effects should be generalizable across neuroimaging modalities, and we should be able to replicate previous findings with fNIRS. The present study therefore addresses whether this temporal neural signature is also feasible in quantifying cognitive load with fNIRS.

1.1. Study design and hypothesis

Based on previous findings in fMRI and EEG studies (Barnes et al., 2009; Churchill et al., 2016; He, 2011, 2014), we hypothesize that increasing levels of cognitive load will be associated with suppression of H , indicating decreased scale invariance as measured by fNIRS signals. To test this in the current study, we examined H in an N-back working memory task while fNIRS data were recorded, in a dataset previously analyzed and reported in Meidenbauer et al. (2021)

The N-back task is a classic working memory paradigm which places high demands on working memory and attention with varying levels of task difficulty and cognitive load (Mencarelli et al., 2019; Owen et al., 2005). In the N-back working memory task, participants are required to match the current word/stimulus with the word presented in the previous Nth trial and make a response. The task's difficulty is adjusted by changing the value of N. In the current study, N is equal to 1, 2, or 3, and defined as ‘1-back’, ‘2-back’ and ‘3-back’ conditions. Here, the ‘1-back’ is the easiest and the ‘3-back’ is the most difficult. Accordingly, we hypothesize that scale invariance of fNIRS signal, measured as Hurst exponent (H), will decline with the increasing task difficulty: whereby the ‘1-back’ will show the highest H , while the ‘3-back’ will have the most suppressed/lowest H , indicating a higher cognitive load. We also hypothesize that H will be higher during rest blocks than during task blocks. We test these hypotheses using H averaged across all frontal channels as well as using a Task partial least squares (Task PLS) multivariate approach which incorporates channels, as we may also see some spatial-specificity in where changes in H are the strongest.

2. Method

This paper uses a dataset previously reported in (Meidenbauer, Choe, Cardenas-Iniguez, Huppert, & Berman, 2021). We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. Participants

Seventy adults participated in this study. Participants were recruited from the Chicago area. Participants were only excluded from participating if they did not have normal or corrected-to-normal visual acuity or reported a history of neurological disorders. 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 approximately 15 min of additional study elements related to a video intervention that were separate from the current work. Our sample size was determined based on hypotheses related to these interventions. The full study procedures lasted between 75 and 90 min. From the original 70 participants, 8 participants were excluded from all data analysis due to technical issues, participant non-compliance with the task procedures, or low-quality data. For this particular analysis, 10 additional participants were excluded due to insufficiently high data quality as defined by the structured noise index (SNI; see section *Exclusion of participants based on SNI*), leading to a final sample of 52 participants. Of the 52 participants analyzed here, 26 were male and 26 were female, and the mean age was 24.5 years ($SD = 6.9$ years). 10 participants identified as Hispanic or Latino/a, 13 as Asian or Asian American, 9 as Black or African American, 13 as White or European American, 5 participants reported 2 or more racial or ethnic identity groups, and 2 preferred not to disclose.

2.2. fNIRS data acquisition

fNIRS data were acquired using a continuous-wave NIRx Nirxport2 device (NIRx Medical Technologies, LLC) and NIRx acquisition software Aurora 1.2 at a sampling rate of ~4.4 Hz. Near-infrared light of two different wavelengths (760 and 850 nm) was used to detect the concentration change of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR). There were 16 sources and 16 detectors, creating 43 channels in total, 33 channels covering the bilateral frontal cortex and 10 channels covering the right parietal cortex. Because of the low quality of parietal data collected, the following analysis focuses on data collected in the frontal region (Fig. 1; See section *Exclusion of noisy channels based on Structured Noise Index*). 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. The AAL2 (Automated Anatomical Labeling; Rolls et al., 2015) parcellation was used to generate the montage and provide as much coverage of the prefrontal cortex (PFC) as possible, including bilateral superior and inferior frontal gyri.

2.3. N-back task

Experimenters first took participants through step-by-step instructions of the N-back task before they began practice. Participants were told that they would see a sequence of short words that are separated by brief fixations (small circles), and

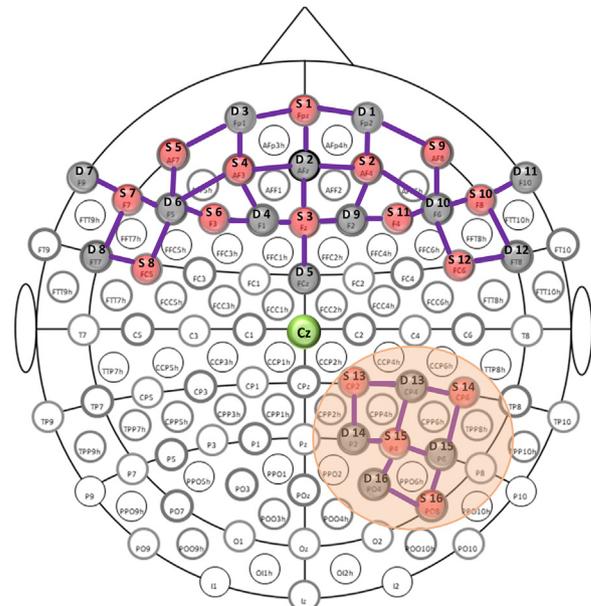


Fig. 1 – Locations of the channels in the International 10/20 coordinates All 33 channels covering bilateral frontal cortex are used in H analysis. 10 Parietal channels (highlighted in orange) were excluded (see Section 2.5.3).

that a word would be presented every 2 sec which should be compared to the word “N” trials back. In the current study, $N = 1, 2, \text{ or } 3$. Participants were instructed to press the “m” key if the current word matched the word N trials back, and to press the “n” key if the current word did not match the word N trials back [Fig. 2]. Blocks began by first displaying the N -back level for that round and a fixation cross (5 sec). Each task block contained a 15-length pseudorandom sequence of two words, presented for 2 sec each (total of 30 sec), followed by 20 sec of rest. The length of each block was 55 sec, and with 18 blocks the total length of the N -back task was approximately 16.5 min. 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 same two words (“AXE” and “BOX”) were used. 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 & Boynton, 2002; Choe et al., 2014; Choe et al., 2016) to suppress autocorrelation. In all cases, words were presented in white text on a black background.

2.4. Study procedures

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. The first round of N -back practice consisted 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

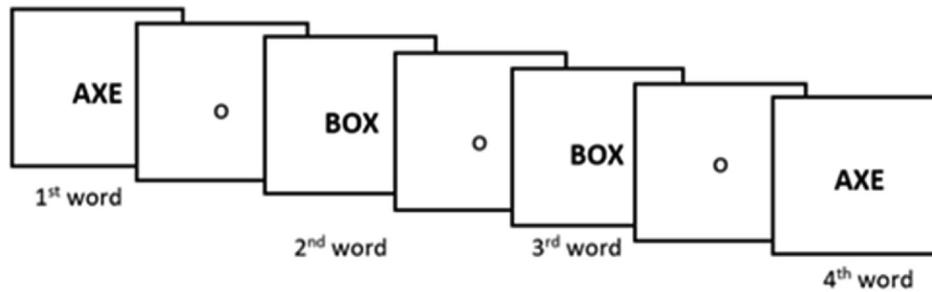


Fig. 2 – The ‘N-back’ task paradigm. Participants were required to respond whether the current word matched the word N trials back (N = 1,2,3). Task difficulty and cognitive load increases with larger N.

of 2-back, then 3 blocks of 3-back, and then one more round of 1-back, 2-back, and 3-back. After the first round of practice, the cap was placed on the participants' 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 calibrated and checked for quality before proceeding. If any channels were not displaying sufficiently high-quality data, placement and hair-clearing were performed again. 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. 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). N-back experiment code can be accessed at <https://osf.io/sh2bf/>

Participants received a performance-based bonus during the main round (18 blocks) of N-back task. The bonus was defined as an additional 40 cents per block if performance >90%, an additional 30 cents per block if performance >80%, and an additional 20 cents per block if performance >60%. 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 sec of task.

2.5. Data analysis

All code can be accessed at: <https://osf.io/kt5cx/>

2.5.1. fNIRS signal preprocessing

fNIRS data were first loaded into the HOMER2 software package (Huppert et al., 2009) for visual inspection of overall data quality (at the level of the participant). This was done by examining the power spectral density plots for all channels to identify the presence of a cardiac oscillation (typically ~1 Hz Tong et al., 2011), which indicates 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. For fNIRS data pre-processing, this study used the Brain AnalyzIR Toolbox (Santosa et al., 2018), and first converted raw light intensity values into

optical density signals. Then, optical density signals were transformed into concentration changes of HbO and HbR values based on the modified Beer–Lambert law (Strangman et al., 2003).

2.5.2. Hurst exponent calculation on HbO and HbR data

To calculate the H value, the key index of interest in this study, we used the Detrended Fluctuations Analysis algorithm, which was developed for fMRI H analysis (DFA) (Churchill et al., 2015, Churchill et al., 2016). We adopted the DFA algorithm to derive H from fNIRS signals in our analysis as fNIRS measures the same biological signal as fMRI.

DFA measures the power in HbO/HbR fluctuations for different time windows of the data, formulated as $F(n)$ as a function of time in windows of length n . The Hurst exponent is equal to the slope α of a linear fit between the log-transformed $F(n)$ and n , with $\alpha = 1$ indicating a perfectly fractal signal. The length of time window, n , in our analysis was set as the full length of rest, which was 20 sec, with which the sampling rate of ~4.4 Hz, yielded 87 samples/block. We kept this time window the same between task and rest so the two could be directly compared. As the task-evoked hemodynamic response occurs on a delay after the onset of the underlying neural activity, the last 20 sec of the 30 sec window of task was used in the DFA calculation.

As each N-back experiment condition had 6 blocks (18 blocks in total), including both task and rest, after the DFA calculation, each participant had 18 H values for task and 18 for rest (6 for each condition) for each of the 33 channels. Further visualization and statistical analysis were based on these H values, the key index of interest.

2.5.3. Exclusion of noisy channels based on structured noise index (SNI)

Before conducting statistical analysis, we first examined data quality based on the Structured Noise Index (SNI), calculated from the Brain AnalyzIR toolbox. SNI was calculated for each channel, for each participant, and is a useful tool in capturing the systematic noise across channels and participants. The SNI is defined as the ratio of the variance of the full data trace to the variance of the auto-regressively whitened trace of the same data and reflects the ratio of structured (colored) noise in the data due to various physiological processes to the uncorrelated (white) noise. This approach was inspired by the spatial SNI method described in Nelson et al. (2014) and applied to the fNIRS time signals in this work. This step was

important as fNIRS signals are sensitive to superficial physiological noise (e.g., hair) and participants have varied levels of interference based on the color and texture of their hair (Santosa et al., 2018). When SNI is less than 2, it indicates that the data from this channel are very noisy and not appropriate for further analysis, and hence, we defined channels with SNI less than 2 as ‘bad’ channels.

Based on this criterion, we found that, out of total 43 channels, participants had on average 6.94 bad channels for HbO (16%) and 9.35 bad channels for HbR (22%), and that these bad channels mainly centered around parietal regions, where signals are more likely to be obstructed by hair. In this dataset, a large number of participants (46 out of 52) had one or more channels marked as “bad” in those covering the right parietal cortex (channels no. 33–43 in Figs. 1 and 3). Therefore, in the following analysis, we excluded the 10 channels in the right parietal cortex, resulting in a total of 33 channels for analysis.

2.5.4. Exclusion of participants based on SNI

Additionally, after excluding those 10 channels in the parietal cortex, 10 participants’ data were found to have many poor-quality channels (shown in Fig. 4). This was defined by participants whose count of bad channels exceeded 1 SD (3.0 bad channels for HbO; 3.4 bad channels for HbR) from the average number of bad channels (2 bad channels for HbO; 3 bad channels for HbR). Following this criterion, participants who either had 5 bad channels for HbO or 6 bad channels for HbR were excluded from further analysis. On average, these 10 outlier participants had 8 (24%) bad channels for HbO and 9 (27%) bad channels for HbR.

As an additional check, after calculating the Hurst exponent, we plotted average H across all task sessions by channels for each participant. As H is a global signal across the whole brain, we expected moderate to high levels of similarity across channels (Churchill et al., 2016; Kardan et al., 2020b). Large fluctuations between channels of the same participant in the same run suggest poor data quality, potentially stemming from hair thickness variations. Shown in Fig. 5, we could see that compared to ‘good fNIRS quality’ participants (upper

panel), ‘poor fNIRS quality’ participants (lower panel) generally had more variable responses across channels that were very irregular and unexpected, which further justified this SNI based participants exclusion step.

2.5.5. Regressing SNI from H values

We further regressed the SNI out from the H value of each channel, which accounts for different degrees of noise across channels, to remove the systematic interference of noisy data. The value of H after SNI being regressed out was used in further analysis.

2.5.6. Accounting for motion-related artifacts

Lastly, as the scale-invariance of a data time series can be influenced by the presence of large fluctuations not due to the underlying biological signal, such as motion-related artifacts (Rubin et al., 2013), we evaluated the presence of such artifacts across conditions and for task and rest separately to ensure these were not driving any results. For example, if participants in the study systematically demonstrated more head movement more during rest blocks than during task blocks, this would lessen the interpretability of our H results. To do this, the raw data were first segmented into the same 20 sec blocks used in the DFA calculation and statistical analysis. Next, statistical outliers in the 20 sec of data were identified by calculating the innovations model for each segment using the function `nirs.math.innovations()` in the Brain AnalyzIR Toolbox (Santosa et al., 2018). The innovations in each time series reflect the uncorrelated (whitened) signal after filtering using the autoregressive model. Here, each 20 sec time series for the 33 channels used in the Hurst analyses for each participant was fed into the function, and the maximum model order was set to 20. This was chosen based on the recommendation that the maximum model order is at least 4 times the sample rate (here ~4.4 Hz).

The output of this function is an innovations time series for each channel for each segment. Subsequently, the number of studentized outliers were calculated (outliers defined as $p < .05$). The counts of these outliers across all 33 channels for

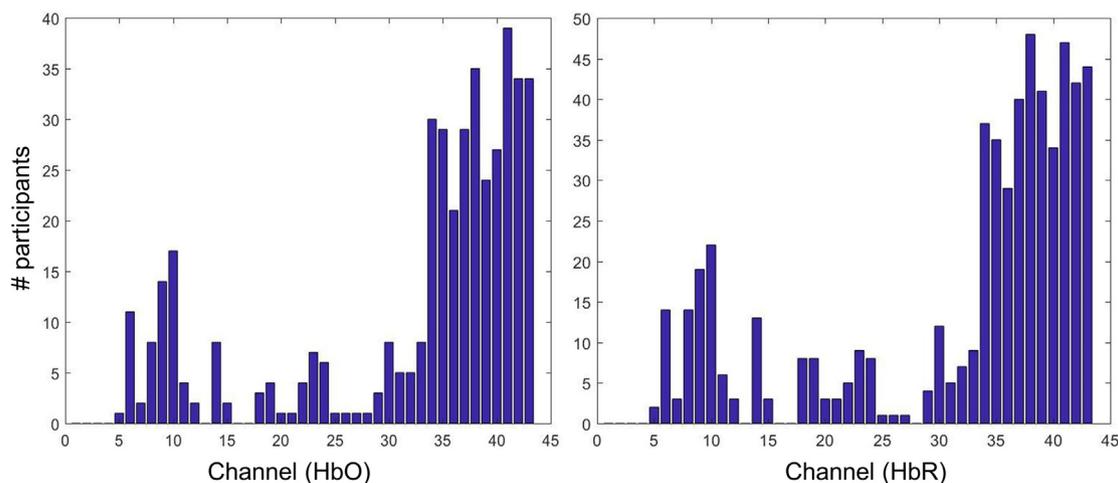


Fig. 3 – Number of participants with each channel (1–43) marked as bad. The number of participants reporting the channel as ‘bad’ were calculated based on channel SNI < 2. Left figure for HbO; Right figure for HbR. Channels 33–43 reflect 10 channels over parietal cortex that were excluded from future analysis.

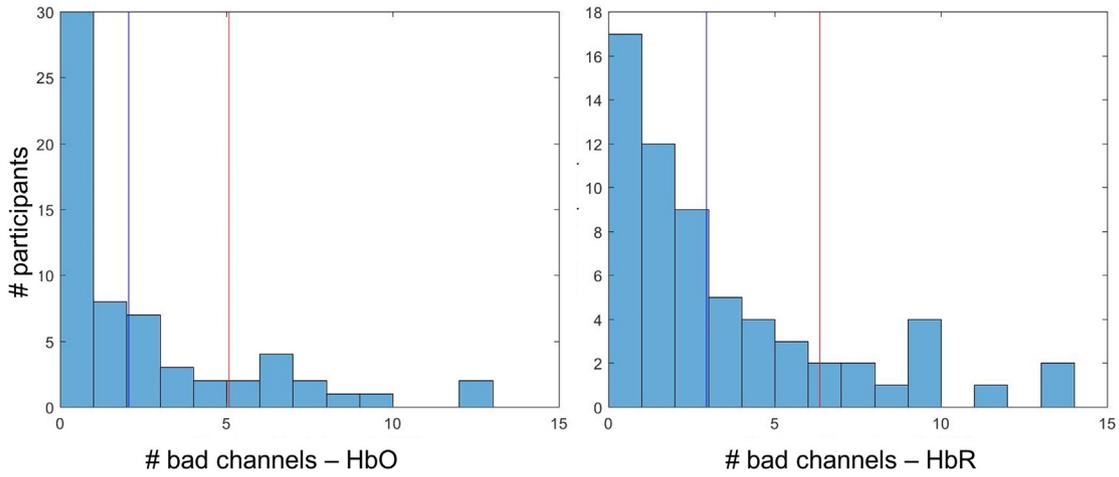


Fig. 4 – Number of ‘bad’ channels per participant after excluding parietal channels. Histograms show the number of ‘bad’ (SNI < 2) channels (out of 33) each participant had after removing the 10 parietal channels for HbO (left) and HbR (right). The blue line represents the mean number of bad channels across participants and the red lines indicate 1 SD from the mean.

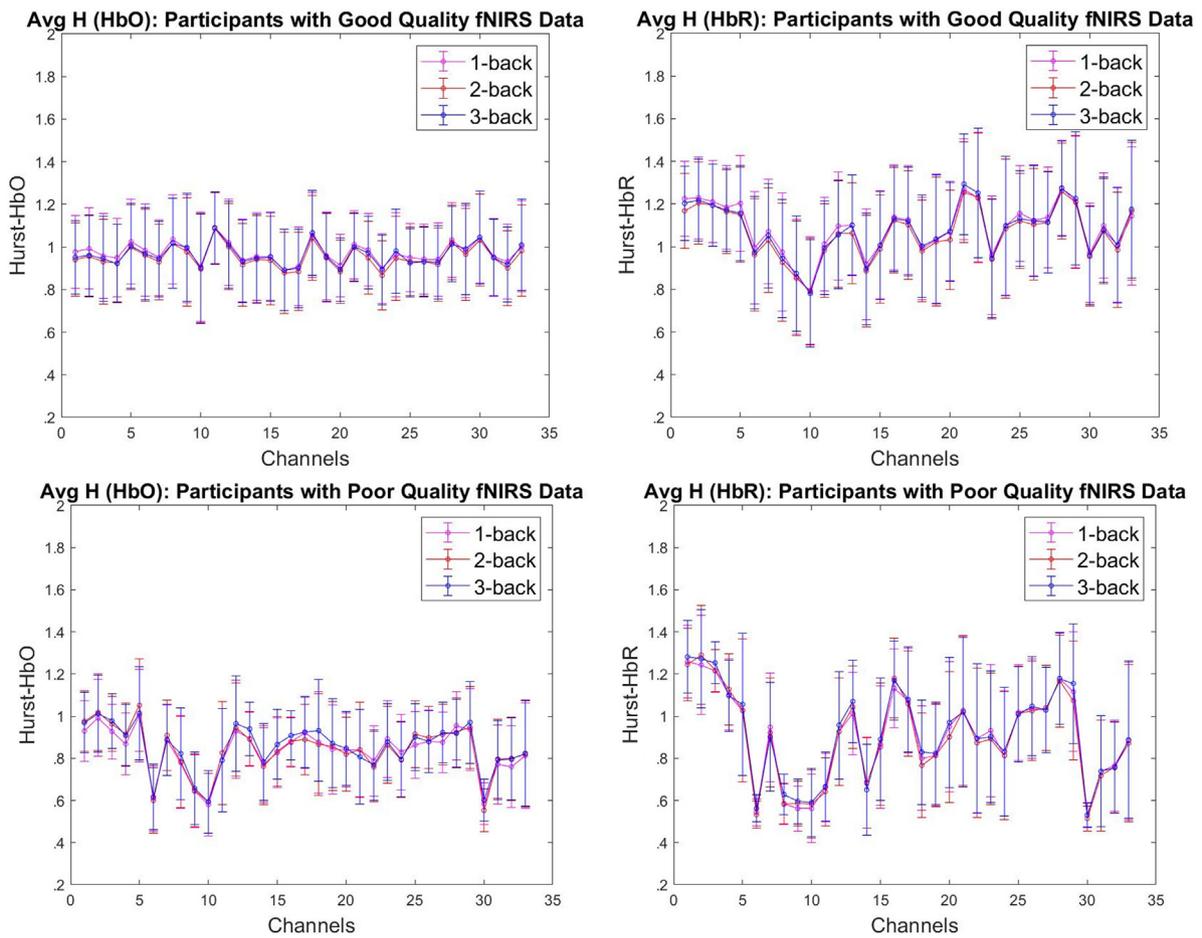


Fig. 5 – Averaged H value by channels for participants with good and poor data quality. H value for each 20 sec segment was averaged for each block type for each participant. The upper figures showed the averaged H for participants marked as having ‘good quality data’ (left for HbO, right for HbR); the lower figures showed the averaged H with error bars reflecting standard deviations for participants which were marked as having ‘poor quality data’ (left for HbO, right for HbR).

each segment and each participant were then saved. The average number of statistical outliers by segment type (all task, all rest, 1-back, 2-back, and 3-back) across are presented in Table 1. Paired t-tests (mimicking the statistical tests on averaged H values, detailed below) were also conducted to compare the number of outliers/artifacts by condition. The results of these tests are also presented in Table 1. The number of outliers was generally quite low: out of ~2870 data points reflecting time (~87 samples) \times 33 channels, the average number of outliers participants had for each condition was between 182 and 200, or roughly 6.3%–6.9% of the data. The counts were also very similar across conditions [Table 1]. For those comparisons which did show a significant difference between conditions, the relative number of outliers were in the opposite direction of what would be expected if these motion artifacts were affecting the H results, as prior work that includes both real data and simulated data all indicate that greater motion artifacts lead to less scale invariant time series and lower H values (Chen et al., 2002; Rubin et al., 2013). Additionally, it is unlikely that this would be driving any effects, as we do not see consistently different motion artifacts across HbO and HbR but similar H effects in both. This suggests that these outliers (which exceed what is a reasonable change to be expected from the biological signal of interest) are not responsible for differences in H values across conditions.

2.5.7. Statistical analysis on average H

Using the cleaned averaged H values with SNI regressed out, planned pairwise t-tests between task and rest were conducted on HbO and HbR separately. Next, two repeated measurements ANOVAs (separately for HbO and HbR) were conducted on average H value with N-back condition (1-back, 2-back, or 3-back) as a within-subjects factor. Lastly, we also examined pairwise t-tests comparing each of the three N-back

conditions. These analyses were carried out using R version 3.5.1 (R Core Team, 2018).

2.5.8. PLS analysis

In addition to statistical analyses on the Hurst values averaged across channels, partial least squares (PLS; Berman et al., 2014; Krishnan et al., 2011; McIntosh & Lobaugh, 2004; <https://www.rotman-baycrest.on.ca>) analyses were also conducted using Matlab v 2018b. PLS is a data-driven, multivariate statistical technique which identifies the relationship between two sets of variables. In neuroimaging research, PLS is often used to find the relationship between neural activity at different spatial locations (e.g., voxels or ROIs in fMRI data, electrodes in EEG data, channels in fNIRS data) and the task design (e.g., experimental conditions or grouping variables). In the current work, a Task PLS was conducted to examine H by N-back level across channels to investigate whether there were specific regions of the PFC that showed a stronger relationship between H and cognitive load and as we may be losing important channel-level information by averaging across the full montage.

The partial least squares analysis relies on the singular value decomposition (SVD) of a covariance matrix. In the Task PLS analysis, the input for SVD is a matrix of the H values for each channel by condition (N-back level) that are averaged across participants (i.e., matrix of 33 channels \times 3 N-back levels). Running an SVD on this 33 \times 3 covariance matrix (R) decomposes it into three matrices: $R = U\Delta V^T$, where the 3 \times 3 matrix U represents the decomposition of R in N-back condition space, the 33 \times 3 matrix V represents the decomposition of R in neural activity channel space, and Δ is the 3 \times 3 diagonal matrix of singular values that quantifies the weighting of each of the singular vectors (i.e., columns in U and V). These linear decompositions which maximize the covariance between brain activity (H values) and task design (N-back level) are referred to as latent variables (LVs). In other words, each LV is comprised of the singular vector V , which reflects a linear combination of channel-level H values (i.e., is a 1 LV \times 33 channel vector) that best characterize R , the singular vector U which reflects the design profile (1 LV \times 3 N-back level vector) that best characterize R , and the weighting of the LV which is represented by a singular value in one column of the diagonal matrix, Δ . These LVs are calculated in order of magnitude of cross-block covariance explained and are mutually orthogonal, so the first latent variable (LV 1) explains the greatest proportion of cross-block covariance, the second latent variable (LV 2) explains the second most proportion, etc.

Ten thousand permutation tests were performed to obtain p -values for each latent variable and 10,000 bootstrapped samples with replacement were created to generate the 95% confidence intervals for variable loadings. The bootstrap ratios (calculated as salience [weights]/standard error [reliability]) measure the reliability of the 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.

Table 1 – Counts & comparisons of average number of motion related artifacts.

Counts	HbO time series M (SD)	HbR time series M (SD)
Task	183.6 (71.8)	197.5 (65.0)
Rest	184.1 (71.6)	198.8 (64.1)
1-back task	182.7 (71.8)	200.1 (67.3)
2-back task	183.1 (72.1)	196.1 (64.4)
3-back task	185.5 (71.8)	196.2 (63.3)
Pairwise T-tests	HbO	HbR
Task versus rest	$t(51) = -.48, p = .63$	$t(51) = -1.78, p = .08$
1-back versus 2-back	$t(51) = -.36, p = .72$	$t(51) = 2.57, p = .01$
2-back versus 3-back	$t(51) = -1.41, p = .16$	$t(51) = -.07, p = .94$
1-back versus 3-back	$t(51) = -1.77, p = .08$	$t(51) = 2.11, p = .04$
The average number of motion artifacts were calculated for each participant in each 20 sec block type (all task, all rest, 1-back, 2-back, and 3-back). Each average count is out of 2871 potential data points (33 channels \times 87 samples). M and SD calculated over the 52 participants' average motion artifacts in each block type. T-tests reflect paired (within-subjects) comparisons of motion artifact counts in each block type.		

3. Results

3.1. Statistical analysis on averaged Hurst exponents

For the task and rest comparison, planned pairwise t-tests with Bonferroni correction (critical $\alpha = .05/2 = .025$) showed that compared with the rest condition, averaged H scores were significantly lower in the task conditions for both HbO (task $M = .83$, $SD = .15$, Rest $M = .96$, $SD = .15$, $t(51) = 11.76$ $p < .001$) and HbR (task $M = .88$, $SD = .17$, Rest $M = .94$, $SD = .16$; $t(51) = 7.72$, $p < .001$). These results are shown in Fig. 6.

For the N-back conditions comparison, where N-back condition is a within-subjects factor, repeated measures ANOVAs on averaged H by both HbO and HbR showed a significant omnibus ANOVA for H extracted from HbR ($F(2,102) = 3.78$, $p = .026$), but not from HbO ($F(2,102) = 1.92$, $p = .153$). In both H extracted from HbO and HbR, 2-back had the lowest value (HbO $M = .82$, $SD = .15$; HbR $M = .86$, $SD = .19$) of H , followed by 3-back (HbO $M = .83$, $SD = .15$; HbR $M = .88$, $SD = .16$) and then 1-back (HbO $M = .84$, $SD = .16$; HbR $M = .89$, $SD = .17$).

For pairwise t-tests of the significant omnibus ANOVA for HbR, with Bonferroni correction (critical $\alpha = .05/3 = .017$), H was significantly different between the 1-back and 2-back conditions ($t(51) = 2.62$, $p = .011$), but not between 2-back and 3-back ($p = .078$) or between 3-back and 1-back ($p = .34$).

3.2. PLS results

Task PLS analyses looking at channel-level H by N-back condition were run separately on H extracted from HbO and from HbR time series. The first latent variable (LV 1) from the analysis with H from deoxyhemoglobin concentrations (HbR) was significant and explained 77% of the crossblock covariance ($p = .005$). LVs 2 and 3 in this analysis were not significant (all $ps > .4$). For the significant LV 1 in H from HbR, 11 mostly medial-frontal channels (#1, #5, #6, #7, #8, #11, #12, #13, #14, #20, and #25) showed stable changes in scale-invariance by N-

back level, indicated by bootstrap ratios with absolute values greater than 2 [Table 2; Fig. 7]. LV 1 from the analysis of H calculated from oxyhemoglobin concentrations (HbO) was not significant ($p = .066$) but did also explain 77% of the crossblock covariance and mimicked the pattern of results found in H from HbR so the results are still presented in Table 2 and Fig. 7. LVs 2 and 3 in this analysis were not significant (all $ps > .17$). For LV 1, 2 channels in the medial superior frontal gyrus (#1 and #2) and 2 in the left inferior frontal gyrus (#22 and #23) showed differences in H by N-back level. In all cases, the significant bootstrap ratios were greater than 2 (and none smaller than -2), indicating the relationship between N-back level and H was in the same direction. In H calculated from HbO and HbR, the first latent variable from each was driven primarily by the contrast between 1-back and 2-back, with higher H found during the 1-back task relative to the 2-back task [Fig. 7]

In addition to the N-back level PLS analyses reported above, supplementary analyses were conducted which included rest blocks. Rest was not included in the primary analyses as there were 3 times as many rest blocks as 1-back, 2-back, and 3-back blocks (as there was a rest after each block), and this would lead to a better signal-to-noise ratio by averaging across more data in rest versus individual N-back level. When rest was included in the PLS model, the first LV examining H extracted from HbO explained 99.3% of the crossblock covariance ($p < .001$), with all 33 channels displaying positive, significant bootstrap ratios >2 . For H extracted from HbR, the first LV explained 95.7% of the crossblock covariance ($p < .001$) with 27 out of 33 channels displaying significant bootstrap ratios (all >2). LV 1 loadings for H from HbO and HbR are shown in Fig. 8 [Top Panel]. The table of significant channels with corresponding ROIs and bootstrap ratios can be found in the Supplementary materials.

Additionally, to further investigate whether the non-linear load effect (H higher for 3-back than 2-back) was due to participants' low accuracy on this task, these same PLS analyses were run separately on the subset of participants ($N = 33$) who scored higher than 80% on average for the 3-back task and the

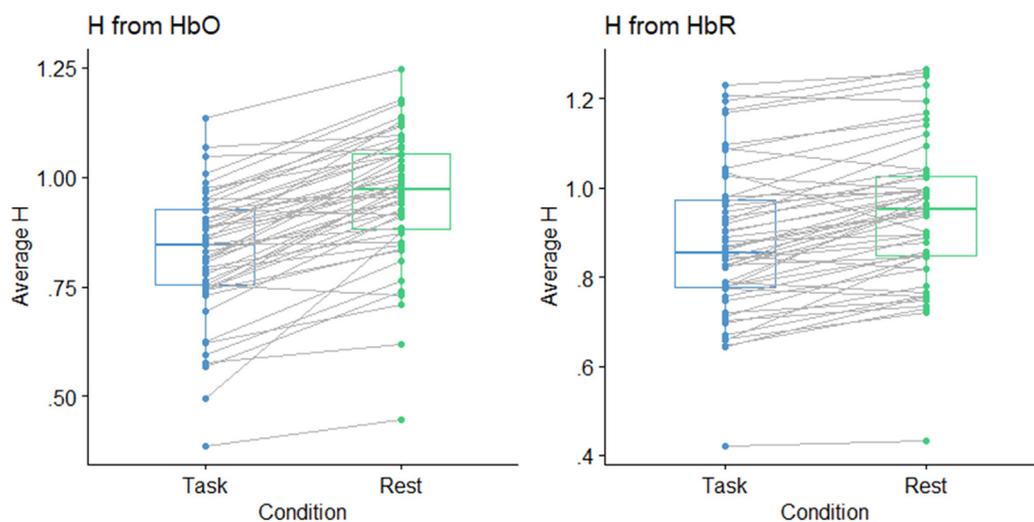


Fig. 6 – Boxplot showing within-subjects effects for averaged H for task versus rest. Gray lines connect H values for rest and task for each participant.

Table 2 – Significant channels for task PLS LV 1 Top: Results for H from deoxyhemoglobin (HbR), bottom: results for H from oxyhemoglobin (HbO). 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
H from HbR				
8	3	2	Medial superior frontal gyrus	4.1
1	1	1	Medial SFG/OFC	4.0
5	2	2	Medial superior frontal gyrus	3.9
6	2	9	Medial superior frontal gyrus	2.9
12	4	2	Medial superior frontal gyrus	2.7
7	2	10	Right inferior frontal gyrus	2.7
14	4	4	Medial superior frontal gyrus	2.5
13	4	3	Medial superior frontal gyrus	2.5
25	9	1	Medial SFG/right IFG	2.4
11	3	9	Medial superior frontal gyrus	2.3
20	7	6	Left inferior frontal gyrus	2.2
H from HbO				
22	7	8	Left inferior frontal gyrus	2.6
2	1	2	Medial superior frontal gyrus	2.6
1	1	1	Medial SFG/OFC	2.4
23	8	6	Left inferior frontal gyrus	2.2

subset ($N = 19$) who scored 80% or lower on average for 3-back. In the high accuracy subset's H from HbO, the first LV explained 95.4% of the crossblock covariance ($p < .001$), with 32 out of 33 channels showing reliable bootstrap ratios (>2). For H from HbR, the first LV explained 81.8% of the crossblock covariance ($p = .008$), with 19 channels generating significant bootstrap ratios (>2) [Fig. 8, Middle Panel]. For the low accuracy subset, the first LV for H from HbO explained 98.4% of the crossblock covariance ($p < .001$), with all 33 channels showing reliable bootstrap ratios (>2). The first LV for H from HbR in this subset explained 91.6% of the crossblock covariance ($p < .001$), with 26 out of 33 channels showing reliable bootstrap ratios (>2) [Fig. 8, Bottom Panel]. Overall, the exclusion of participants with poor accuracy on 3-back led to a much clearer effect of N-back level on H in the expected direction. The table of bootstrap ratios with ROIs are reported in the [Supplementary materials](#).

4. Discussion

Previous research suggests that when people are performing a cognitively demanding task, the temporal property of their brain signals will be less scale invariant than at rest, as quantified by lower H . In addition, this suppression of scale invariance is found across the whole brain (He, 2011, 2014; Kardan et al., 2020a) and is unidirectional (Churchill et al., 2016). However, whether this signature could be extracted from fNIRS data was unclear. The present study is the first to apply a scale-invariance (Hurst exponent, H) analysis to measuring cognitive load with fNIRS, which can be used in a much wider variety of settings than other neuroimaging modalities. Consistent with previous neuroimaging research, we found that task and rest conditions significantly differed by their average H calculated from both oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) concentrations changes of the fNIRS signal. Compared with rest, average H for task was significantly lower, which suggests a higher level of cognitive

effort and difficulty while performing the N-back task relative to rest. For N-back condition, a more subtle manipulation of cognitive load, pairwise t-tests also showed a significant difference between 1-back and 2-back by H calculated from HbR. This was not significant with H from HbO, where again, H during 2-back had lower values than 1-back, but they do demonstrate convergence with the pattern observed in HbR and in comparing task vs rest.

The N-back level results from averaging across all 33 channels were further supported by those yielded from the partial least squares analysis. Specifically, in the Task PLS, which examines differences in H across channels and N-back level, the first latent variable from HbR demonstrated a robust effect of higher H in 1-back versus 2-back across 11 frontal channels. Though the effect was weaker in H calculated from HbO, the same pattern was found in this analysis' LV 1. These results provide complementary evidence for an effect of cognitive load and task difficulty on H derived from fNIRS. Additionally, they suggest that while H is believed to be a relatively global brain signature (Churchill et al., 2016), greater sensitivity can be achieved by adopting a multivariate technique which takes the full data (including channels) as input.

It is worth noting that both in the pairwise comparisons of average H by N-back condition and the Task PLS analysis, this study only yielded significant results for the 1-back versus 2-back comparison. In contrast, the 2-back versus 3-back comparison was marginally significant for H by HbR (not with HbO), and all the other t-test results for 3-back condition (3-back vs 1-back and 3-back vs 2-back) were not significant. Moreover, as observed in Meidenbauer et al. (2021) which used this same dataset, performance in 3-back condition was generally low and was highly variable. The non-significant results for the 3-back condition have also been shown in previous studies, indicated by a non-linear effect of N-back load (Aghajani et al., 2017; Mandrick et al., 2013, 2016). Researchers have argued that if a task is too difficult, people may disengage from it or simply "give up", since it exceeds one's capability (Causse et al., 2017; Luctkar-Flude and Groll, 2015b;

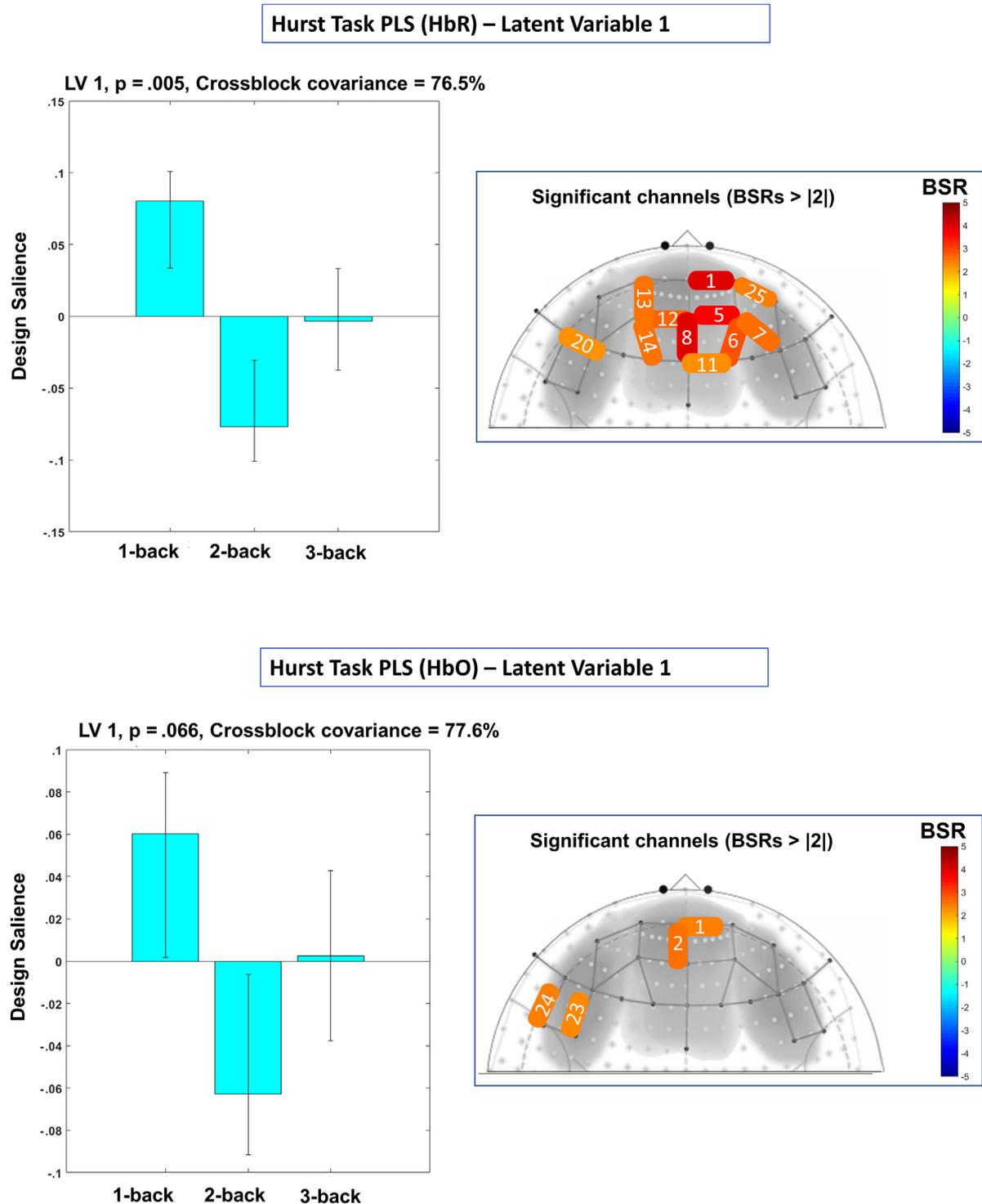


Fig. 7 – LV 1 demonstrated an N-back load-dependent relationship with H extracted from deoxyhemoglobin (HbR; top panel) and oxyhemoglobin (HbO; bottom panel) concentrations. The left side plots show the relation between H and N-back level. Error bars are 95% confidence intervals around the mean design saliency value. The right panel shows channels (labeled by number), which had bootstrap ratios (BSR) > |2|.

Mandrick et al., 2013, 2016). Following this argument and based on the results reported in Meidenbauer et al. (2021), we infer that in this study, the 3-back condition might not be reflecting the highest cognitive load across all participants. Supporting this idea further, supplementary PLS analyses

excluding participants with worse than 80% accuracy on the 3-back task did show the expected load effect in H extracted from HbR and to a lesser extent, in H from HbO. However, future work is needed to evaluate the extent to which H extracted from fNIRS is affected by individual differences in

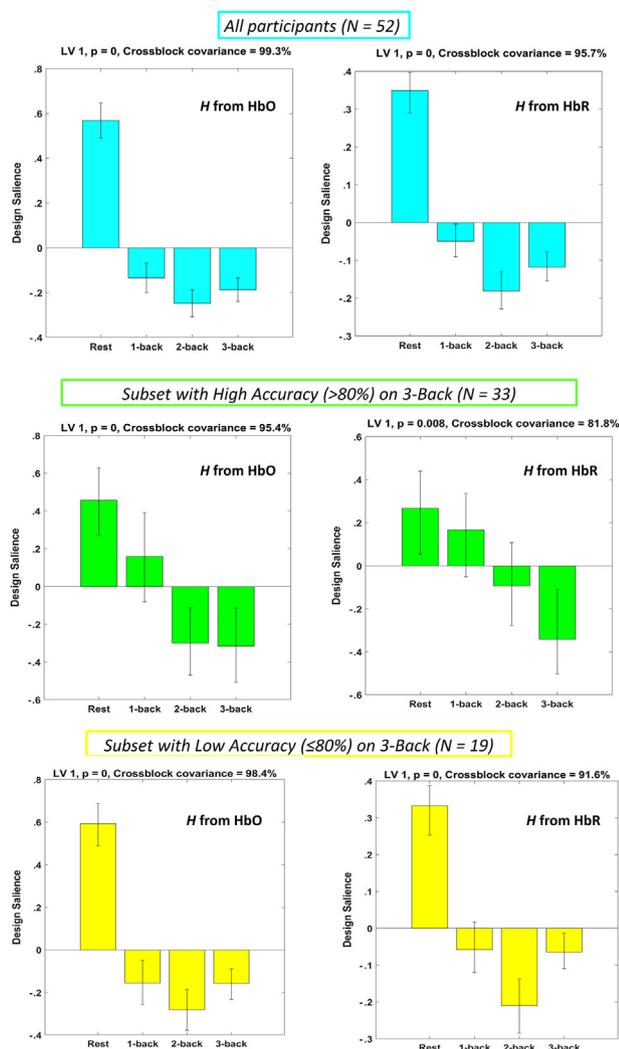


Fig. 8 – LV 1 examining H as a function of N -back level and rest. Plots show the relation between H and Rest versus each N -back level. Error bars are 95% confidence intervals around the mean design saliency value. **TOP PANEL:** Including all participants ($N = 52$). All channels for H from HbO and 27 out of 33 channels for HbR showed significant, positive bootstrap ratios ($BSR > 2$). **MIDDLE PANEL:** Only participants scoring higher than 80% on the 3-back task ($N = 33$). In H from HbO, 32 out of 33 channels had $BSRs > 2$. For H from HbR, 19 out of 33 channels had $BSRs > 2$. **BOTTOM PANEL:** Only participants scoring 80% or lower on the 3-back task ($N = 19$). In H from HbO, 33 channels had $BSRs > 2$. In H from HbR, 26 channels had $BSRs > 2$. No bootstrap ratios were < -2 in any analyses.

effort and task disengagement relative to explicitly manipulated cognitive load.

4.1. Data preprocessing pipeline for H analysis with fNIRS

The present work involved developing a basic pipeline for data preprocessing and H analysis with fNIRS (All analysis scripts

can be accessed at: <https://osf.io/kt5cx/>). To check the quality of our data, we used the SNI (Structured Noise Index) to measure the systematic noise across channels. We subsequently excluded participants who had a high number of low SNI channels and regressed out SNI from the Hurst exponent. Both steps were supported by visualizations as effective in detecting poor data quality, finding 'outliers', and removing systematic noise. These steps were adopted for two primary reasons. First, superficial physiological noise is structured in fNIRS and can differ by person according to how their hair may obstruct signals (Huppert, 2016; Santosa et al., 2018). Second, scale invariance is generally a whole brain index that, when examined across channels, could be biased by channel-level differences in superficial noise. To further rule out the influence of motion-related artifacts, we used an innovations model to calculate the statistical outliers of the time series (i.e., the signal variations which are larger than what is expected due to underlying physiological changes) and found that the motion-related artifacts did not explain H effects.

4.2. Implications of the current study

This study sheds light on the reliability of scale invariance across neuroimaging modalities and on the promising future of adopting fNIRS in examining cognitive load in real life scenarios. Relative to EEG and fMRI, fNIRS is more flexible, less affected by environmental noise, and more robust to motion artifacts. fNIRS is already used to study cognitive processes in more ecological valid scenarios (Liu et al., 2016; Quaresima & Ferrari, 2019). The results of this study demonstrate the effectiveness of H in measuring cognitive load with fNIRS and further strengthens its capability in real-world settings, such as monitoring cognitive load during driving, social interactions, or even to examine cognitive restoration during real interaction with natural versus urban environments (Berman et al., 2008).

4.3. Limitations and future directions

As this study is the first to demonstrate the effectiveness of H with fNIRS in measuring cognitive load, there are several limitations which require further investigation. First, applying scale invariance analysis in various datasets and experiment settings (especially in real-world scenarios) would be necessary in the future to further validate the effectiveness and robustness of this measure of cognitive load with fNIRS. Secondly, since this analysis measures scale invariance on a temporal scale, the time window of analysis might impact the results. In this study, we adopted 20 sec as the time window to match the full length of the rest session. As fNIRS has a relatively high sampling rate, 20 sec \times 4.5 Hz provided a sufficiently high number of samples for the DFA to be reliable. Additionally, due to the high correspondence between fNIRS and fMRI, we adopted the DFA algorithm from fMRI Hurst exponent analysis (Churchill et al., 2016) and showed its effectiveness with the current fNIRS pre-processing pipeline. However, future work should explore other non-stationary algorithms to calculate H and examine whether and how different time window lengths might impact its effectiveness.

We found that H calculated from deoxyhemoglobin (HbR) showed more reliable effects than did H from oxyhemoglobin

(HbO). This may be due to the fact that HbR is more tightly coupled with the BOLD response in fMRI (Huppert et al., 2006). However, as this is the first study to look at H in fNIRS, it is unclear whether this is a reliable pattern or is related to the current task design. Understanding why H extracted from HbR shows a stronger effect would be an important and exciting future direction for this work. Ideally, future investigations could employ different tasks, contexts, and time windows to better illuminate these possibilities.

While the current dataset was well-suited to conduct an initial validation of H from fNIRS as a measure of cognitive load as it involved a large sample and a standardized working memory task, it is worth noting that sequential analysis of the same dataset has the potential to increase the Type-I error rate (Thompson et al., 2020). Thus, while reanalyzing this existing dataset provides a useful first step, future research is needed to more thoroughly validate the robustness of these results in other data sources. Lastly, as the sample examined in the current study was primarily young adults (~24.5 years old on average), we were not able to examine the effects of age on H . Research in fMRI has identified that younger adults tend to have higher H than older adults, and whether this age-effect can be replicated in fNIRS remains an open question.

5. Conclusion

This study validated the Hurst exponent as an effective measure of cognitive load with fNIRS, opening the door for a wide variety of applications for monitoring cognitive load and fatigue in ecologically valid settings. This work demonstrated a basic and robust pipeline for calculating scale-invariance analysis in fNIRS and lays the foundation for future theoretical and practical research using this method. Future work could further test its theoretical validity and explore its implications with fNIRS in the real world.

Data and code availability statement

Raw data, processed data, H calculation code, and all statistical analysis code can be accessed at the OSF project page: <https://osf.io/kt5cx/>

PsychoPy N-back Experiment code can be accessed at the OSF page associated with Meidenbauer et al., *NeuroImage* (2021): <https://osf.io/sh2bf/>.

Study procedures and analyses were not pre-registered prior to the research being conducted.

Funding

Supported by the National Science Foundation BCS-1632445 and DRL-2055420 to M.G.B. and the Mansueto Institute for Urban Innovation Postdoctoral Fellowship to K.W.C.

Our fNIRS device was provided by the University of Chicago Neuroscience Institute Shared Equipment Award.

Author CRediT statement

Conceptualization: Kimberly L. Meidenbauer, Kyoung W. Choe, and Marc G. Berman.

Data curation: Chu Zhuang and Kimberly L. Meidenbauer.

Formal analysis: Chu Zhuang, Kimberly L. Meidenbauer, Omid Kardan, Andrew J. Stier, Kyoung W. Choe, Theodore J. Huppert, and Marc G. Berman.

Funding acquisition: Kyoung W. Choe and Marc G. Berman.

Investigation: Kimberly L. Meidenbauer and Kyoung W. Choe.

Methodology: Kyoung W. Choe and Theodore J. Huppert.

Project administration: Kimberly L. Meidenbauer, Kyoung W. Choe, and Marc G. Berman.

Resources: Omid Kardan, Andrew J. Stier, Kyoung W. Choe, Carlos Cardenas-Iniguez, and Theodore J. Huppert.

Software: Chu Zhuang, Kimberly L. Meidenbauer, Kyoung W. Choe, Carlos Cardenas-Iniguez, and Theodore J. Huppert.

Supervision: Marc G. Berman.

Visualization: Chu Zhuang and Kimberly L. Meidenbauer.

Writing – original draft: Chu Zhuang and Kimberly L. Meidenbauer.

Writing – review & editing: Chu Zhuang, Kimberly L. Meidenbauer, Omid Kardan, Andrew J. Stier, Kyoung W. Choe, Theodore J. Huppert, and Marc G. Berman.

Open practices

The study in this article earned Open Data and Open Materials badges for transparent practices. Materials and data for the study are available at <https://osf.io/kt5cx/>.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

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

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cortex.2022.05.009>.

REFERENCES

Aghajani, H., Garbey, M., & Omurtag, A. (2017). Measuring mental workload with EEG+fNIRS. *Frontiers in Human Neuroscience*, 11, 359.

- Barnes, A., Bullmore, E. T., & Suckling, J. (2009). Endogenous human brain dynamics recover slowly following cognitive effort. *PLoS One*, 4, e6626.
- Berman, M. G., Jonides, J., & Kaplan, S. (2008). The cognitive benefits of interacting with nature. *Psychological Science*, 19. Available <http://journals.sagepub.com/doi/pdf/10.1111/j.1467-9280.2008.02225.x>.
- Berman, M. G., Misic, B., Buschkuhl, M., Kross, E., Deldin, P. J., Peltier, S., ... Jonides, J. (2014). Does resting-state connectivity reflect depressive rumination? A tale of two analyses. *NeuroImage*, 103, 267–279.
- Boedecker, J., Obst, O., Lizier, J. T., Mayer, N. M., & Asada, M. (2012). Information processing in echo state networks at the edge of chaos. *Theory in Biosciences*, 131, 205–213.
- Buracas, G. T., & Boynton, G. M. (2002). Efficient design of event-related fMRI experiments using M-sequences. *NeuroImage*, 16, 801–813.
- Cause, M., Chua, Z., Peysakhovich, V., Del Campo, N., & Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Scientific Reports*, 7, 5222.
- Chen, Z., Ivanov, P. C., Hu, K., & Stanley, H. E. (2002). Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 65, 41107.
- Choe, K. W., Blake, R., & Lee, S.-H. (2014). Dissociation between neural signatures of stimulus and choice in population activity of human V1 during perceptual decision-making. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, 34, 2725–2743.
- Choe, K. W., Blake, R., & Lee, S.-H. (2016). Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vision Research*, 118, 48–59.
- Churchill, N. W., Cimprich, B., Askren, M. K., Reuter-Lorenz, P. A., Jung, M. S., Peltier, S., ... Berman, M. G. (2015). Scale-free brain dynamics under physical and psychological distress: Pre-treatment effects in women diagnosed with breast cancer. *Human Brain Mapping*, 36, 1077–1092.
- Churchill, N. W., Spring, R., Grady, C., Cimprich, B., Askren, M. K., Reuter-Lorenz, P. A., ... Berman, M. G. (2016). The suppression of scale-free fMRI brain dynamics across three different sources of effort: Aging, task novelty and task difficulty. *Scientific Reports*, 6, 30895.
- Cocchi, L., Gollo, L. L., Zalesky, A., & Breakspear, M. (2017). Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in Neurobiology*, 158, 132–152.
- R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available <https://www.R-project.org/>.
- Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*, 54, 2808–2821.
- Dehghani, N., Bédard, C., Cash, S. S., Halgren, E., & Destexhe, A. (2010). Comparative power spectral analysis of simultaneous electroencephalographic and magnetoencephalographic recordings in humans suggests non-resistive extracellular media. *Journal of Computational Neuroscience*, 29, 405–421.
- Eke, A., Herman, P., Kocsis, L., & Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, 23, R1–R38.
- Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63, 921–935.
- Fishburn, F. A., Norr, M. E., Medvedev, A. V., & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience*, 8, 76.
- Freeman, W. J., Holmes, M. D., Burke, B. C., & Vanhatalo, S. (2003). Spatial spectra of scalp EEG and EMG from awake humans. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 114, 1053–1068.
- Freeman, W. J., & Zhai, J. (2009). Simulated power spectral density (PSD) of background electrocorticogram (ECoG). *Cognitive Neurodynamics*, 3, 97–103.
- Garbarino, E. C., & Edell, J. A. (1997). Cognitive effort, affect, and choice. *Journal of Consumer Research*, 24, 147–158.
- Gautam, S. H., Hoang, T. T., McClanahan, K., Grady, S. K., & Shew, W. L. (2015). Maximizing sensory dynamic range by tuning the cortical state to criticality. *PLOS Computational Biology*, 11, e1004576.
- Grady, C. L., Protzner, A. B., Kovacevic, N., Strother, S. C., Afshin-Pour, B., Wojtowicz, M., ... McIntosh, A. R. (2010). A multivariate analysis of age-related differences in default mode and task-positive networks across multiple cognitive domains. *Cerebral Cortex*, 20, 1432–1447.
- He, B. J. (2011). Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, 31, 13786–13795.
- He, B. J. (2014). Scale-free brain activity: Past, present, and future. *Trends in Cognitive Sciences*, 18, 480–487.
- He, B. J., Zempel, J. M., Snyder, A. Z., & Raichle, M. E. (2010). The temporal structures and functional significance of scale-free brain activity. *Neuron*, 66, 353–369.
- Hocke, L. M., Oni, I. K., Duszynski, C. C., Corrigan, A. V., Frederick, B. D., & Dunn, J. F. (2018). Automated processing of fNIRS data—A visual guide to the pitfalls and consequences. *Algorithms*, 11. <https://doi.org/10.3390/a11050067>
- Huppert, T. J. (2016). Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics*, 3, 10401.
- Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, 48, D280–D298.
- Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, 29, 368–382.
- Kardan, O., Adam, K. C. S., Mance, I., Churchill, N. W., Vogel, E. K., & Berman, M. G. (2020a). Distinguishing cognitive effort and working memory load using scale-invariance and alpha suppression in EEG. *NeuroImage*, 211, 116622.
- Kardan, O., Layden, E., Choe, K. W., Lyu, M., Zhang, X., Beilock, S. L., Rosenberg, M. D., Berman, M. G., et al. (2020b). Scale-invariance in brain activity predicts practice effects in cognitive performance. *bioRxiv*. <https://doi.org/10.1101/2020.05.25.114959>
- Kennedy, D. O., & Scholey, A. B. (2000). Glucose administration, heart rate and cognitive performance: Effects of increasing mental effort. *Psychopharmacology*, 149, 63–71.
- Kinouchi, O., & Copelli, M. (2006). Optimal dynamical range of excitable networks at criticality. *Nature Physics*, 2, 348–351.
- Klaassen, E. B., Evers, E. A. T., de Groot, R. H. M., Backes, W. H., Veltman, D. J., & Jolles, J. (2014). Working memory in middle-aged males: Age-related brain activation changes and cognitive fatigue effects. *Biological Psychology*, 96, 134–143.
- Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The twenty electrode system of the international federation. *The international federation of clinical neurophysiology. The international federation of clinical neurophysiology. Supplement, Electroencephalography and Clinical Neurophysiology. Supplement*, 52, 3–6.

- Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial least squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*, 56, 455–475.
- Liu, T., Pelowski, M., Pang, C., Zhou, Y., & Cai, J. (2016). Near-infrared spectroscopy as a tool for driving research. *Ergonomics*, 59, 368–379.
- Lloyd-Fox, S., Blasi, A., & Ellwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience and Biobehavioral Reviews*, 34, 269–284.
- Luctkar-Flude, M., & Groll, D. (2015a). A systematic review of the safety and effect of neurofeedback on fatigue and cognition. *Integrative Cancer Therapies*, 14, 318–340.
- Luctkar-Flude and Groll, 2015b. Luctkar-Flude, M., & Groll, D. (2015). A systematic review of the safety and effect of neurofeedback on fatigue and cognition. *Integrative Cancer Therapies*, 14(4), 318–340.
- Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10, 422–437.
- 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. *Neuroscience Research*, 76, 156–162.
- Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology*, 121, 62–73.
- Maxim, V., Sendur, L., Fadili, J., Suckling, J., Gould, R., Howard, R., & Bullmore, E. (2005). Fractional Gaussian noise, functional MRI and Alzheimer's disease. *NeuroImage*, 25, 141–158.
- McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: Applications and advances. *NeuroImage*, 23(Suppl. 1), S250–S263.
- McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of Cognitive Neuroscience*, 15, 394–408.
- Meidenbauer, K. L., Choe, K. W., Cardenas-Iniguez, C., Huppert, T. J., & Berman, M. G. (2021). Load-dependent relationships between frontal fNIRS activity and performance: A data-driven PLS approach. *NeuroImage*, 230, 117795.
- 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. *Human Brain Mapping*, 40, 3810–3831.
- Miller, K. J., Sorensen, L. B., Ojemann, J. G., & den Nijs, M. (2009). Power-law scaling in the brain surface electric potential. *PLoS Computational Biology*, 5, e1000609.
- Morais, G. A. Z., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decoder (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports*, 8, 3341.
- Naccache, L., Dehaene, S., Cohen, L., Habert, M.-O., Guichart-Gomez, E., Galanaud, D., & Willer, J. C. (2005). Effortless control: Executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia*, 43, 1318–1328.
- Nelson, J. S., Christianson, O. I., Harkness, B. A., Madsen, M. T., Mah, E., Thomas, S. R., ... Samei, E. (2014). Improved nuclear medicine uniformity assessment with noise texture analysis. *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine*, 55, 169–174.
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25, 46–59.
- Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, 35, 737–743.
- Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & Behavior*, 93, 369–378.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203.
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2018). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*. Available <https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/nyas.13948>.
- Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: A concise review. *Organizational Research Methods*, 22, 46–68.
- Rolls, E. T., 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.
- Rubin, D., Fekete, T., & Mujica-Parodi, L. R. (2013). Optimizing complexity measures for fMRI data: Algorithm, artifact, and sensitivity. *PLoS One*, 8, e63448.
- Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS brain AnalyzIR toolbox. *Algorithms*, 11, 73.
- Sievertsen, H. H., Gino, F., & Piovesan, M. (2016). Cognitive fatigue influences students' performance on standardized tests. *Proceedings of the National Academy of Sciences*, 2621–2624. <https://doi.org/10.1073/pnas.1516947113>
- Soltanlou, M., Sitnikova, M. A., Nuerk, H.-C., & Dresler, T. (2018). Applications of functional near-infrared spectroscopy (fNIRS) in studying cognitive development: The case of mathematics and language. *Frontiers in Psychology*, 9, 277.
- Stier, A. J., Cardenas-Iniguez, C., Kardan, O., Moore, T. M., Meyer, F. A. C., Rosenberg, M. D., ... Berman, M. G. (2021). A scale-free gradient of cognitive resource disruptions in childhood psychopathology. *bioRxiv*. <https://doi.org/10.1101/2021.08.24.457554>
- Strangman, G., Franceschini, M. A., & Boas, D. A. (2003). Factors affecting the accuracy of near-infrared spectroscopy concentration calculations for focal changes in oxygenation parameters. *NeuroImage*, 18, 865–879.
- Thompson, W. H., Wright, J., Bissett, P. G., & Poldrack, R. A. (2020). Dataset decay and the problem of sequential analyses on open datasets. *Elife*, 9. <https://doi.org/10.7554/eLife.53498>
- Tong, Y., Lindsey, K. P., & deB Frederick, B. (2011). Partitioning of physiological noise signals in the brain with concurrent near-infrared spectroscopy and fMRI. *Journal of Cerebral Blood Flow & Metabolism*, 31, 2352–2362.