Multi-timepoint pattern analysis (MTPA):

Improving classification with neural timeseries data

Bear M. Goldstein[1], Agnieszka Pluta[2], Grace Q. Miao[1], Ashley L. Binnquist[1], and Matthew D. Lieberman[1]

[1] Department of Psychology, University of California, Los Angeles

[2] Department of Psychology, University of Warsaw, Warsaw, Poland

Author Note

Correspondence concerning this article should be addressed to Bear Goldstein (beargoldstein@ucla.edu) or Matthew Lieberman (lieber@ucla.edu), University of California, Los Angeles, Department of Psychology, 1285 Psychology Building Box 951563, Los Angeles, CA 90095-1563.

© The Author(s) 2025. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Abstract

Long, naturalistic stimuli are effective in evoking meaningfully differential neural response patterns between groups. However, the resulting timeseries data often have a high number of features compared to a limited sample size, increasing the likelihood of overfitting and reducing predictive power. This paper introduces multi-timepoint pattern analysis (MTPA) as a temporal dimension reduction approach for improving prediction accuracy when building models with long neural timeseries data. Using feature selection with elastic net regression, MTPA identifies predictive neural patterns while preserving the temporal structure and interpretability of the data. Across two experiments with distinct populations and objectives, MTPA demonstrated consistent advantages over approaches using principal component analysis, windowed averaging, and no dimension reduction. Experiment 1 predicted persistent work-related psychological states in business professionals, achieving accuracies up to 79.1%. Experiment 2 predicted cognitive load and narrative context during video viewing in undergraduates, with accuracies up to 66.5%. These findings suggest that MTPA may be a useful tool for analyzing neural data from extended naturalistic designs, enabling researchers to improve prediction accuracy across diverse outcomes and obtain new insights into the temporal dynamics of neural responses.

Keywords: machine learning, neural synchrony, fNIRS, prediction, feature selection, timeseries

Word count: 5,127

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

It is widely believed that a 10:1 sample-to-feature ratio is optimal for machine learning classification (Concato et al., 1995; Peduzzi et al., 1996; c.f., Vittinghoff & McCulloch, 2007). This poses a challenge for neuroscientists, as the cost, time, and resource-intensive nature of neuroimaging typically constrain experimental sample size. At the same time, naturalistic stimuli, such as movie clips, are becoming more common in brain-as-predictor approaches (e.g., Dieffenbach et al., 2021; Li et al., 2022; Parkinson et al., 2018; Yeshurun et al., 2017). While these stimuli are beneficial for evoking distinguishable neural response patterns between groups that extend to real-world processes (Jääskeläinen et al., 2021; Nastase et al., 2020; Finn et al., 2020), they necessarily generate a high number of features. A typical study might include 30 participants (samples) but hundreds of timepoints (features), creating a sample-to-feature ratio closer to 1:10 instead of 10:1. This can increase the risk of overfitting and decrease models' ability to predict out-of-sample cases (Guyon & Elisseeff, 2003; Hua et al., 2009; Kohavi & John, 1997). To retain the richness of extended stimuli while still enabling robust predictive modeling, we propose reducing the temporal dimensionality of neural timeseries data prior to classification.

Neuroscientists currently address the problem of high dimensionality in a spatial context by performing multivoxel pattern analysis (MVPA). In some cases, fMRI may record 100,000 voxels per volume but only a few hundred or so trials (Mwangi et al., 2014), so MVPA applies dimension reduction to remove uninformative voxels prior to training the classifier (Weaverdyck et al., 2020). No such systematic approach exists for social neuroscientists aiming to reduce the dimensionality of a long neural timeseries following a naturalistic stimulus, so we introduce *multi-timepoint pattern analysis* (MTPA) as a temporal analog to MVPA. While MVPA

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

performs a spatial searchlight to select for the most predictive voxels, channels, or regions, MTPA exercises feature selection over a timeseries, effectively serving as a temporal searchlight to determine the most predictive timepoints.

A promising approach within MTPA is feature selection through elastic net regression, which combines penalties from lasso and ridge regression to reduce dimensionality while retaining the most predictive features (Zou & Hastie, 2005). By selecting for individual features without transforming or aggregating the data, elastic net preserves interpretability between the features, cognitive processes, and timepoints in the stimulus. Elastic net-based MTPA was first developed for another study (Goldstein et al., 2025) to predict psychological outcomes in business professionals. While this initial work demonstrated the method's utility, it focused exclusively on elastic net and did not compare it to other dimension reduction approaches, including feature extraction techniques like principal component analysis (PCA; Jolliffe & Cadima, 2016) or down-sampling techniques like windowed-averaging.

Present study

The purpose of this study is to explore and validate the MTPA approach, offering a new tool for building predictive models with long neural timeseries. To show how this method can be applied generally and affect prediction accuracy for a range of outcomes, we apply MTPA to two separate datasets, each with different scientific objectives, populations, and outcome measures.

Experiment 1 aims to predict whether real-world businesspeople feel overwhelmed, burned out, or in need of a new or different challenge in their career. Experiment 2 aims to predict a) whether people are experiencing cognitive load while they watch a video and b)

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

whether people are given context for the protagonist's behavior prior to watching the video. Together, these experiments allow us to explore how different approaches perform when predicting persistent yet malleable mental states (Experiment 1), cognitive processes (Experiment 2), and experimentally-induced narrative perspectives (Experiment 2). Across both experiments, participants' brain responses were measured with functional near infrared spectroscopy (fNIRS), which serves as an especially useful tool for predicting with extended naturalistic stimuli given its superior portability, tolerance for movement, and participant comfort compared to fMRI and EEG (Burns & Lieberman, 2019; Scholkmann et al., 2014).

While MTPA uses feature selection to reduce dimensionality, we also compare its performance to other dimension reduction techniques, including a feature extraction-based MTPA using PCA, down-sampling with windowed averages, and models with no temporal dimension reduction. We focus our analyses on neural activity in the dorsal mPFC (dmPFC), anterior mPFC (amPFC), and TPJ – fNIRS-accessible default mode network hubs respectively implicated in mentalizing and social cognition, affective and self-referential processing, and the construction of subjective construals (Lieberman, 2022; Lieberman et al., 2019; Roy et al., 2012).

Across both studies, we expect that MTPA will outperform the other approaches in prediction accuracy. Although we expect feature extraction-based MTPA to predict outcomes more accurately than models that use down-sampling and no dimension reduction, we expect feature selection-based MTPA to be most generally effective, ultimately making the case for an interpretable and predictive approach tailored to the temporal structure of extended naturalistic stimuli.

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

While the main scope of this project is to compare MTPA to more conventional analytic approaches, we also have hypotheses surrounding the predictive capability of each ROI. For predicting people's psychological lenses surrounding their careers in Experiment 1 and whether people were given narrative context in Experiment 2, we expect that all ROIs will be predictive in some capacity given their roles in high-level meaning making and socio-cognitive processing broadly. For Experiment 2, we expect that cognitive load will affect activity in medial prefrontal areas while TPJ processing will remain unchanged (see the CEEING Model of Pre-Reflective Subjective Construal; Lieberman, 2022), so only activity in the amPFC and dmPFC should reliably predict whether people are experiencing cognitive load.

Experiment 1: Methods

The data for Experiment 1 were reported in a separate study using a single analytical approach and focusing on the psychological implications of the findings (Goldstein et al., 2025). That paper provides a comprehensive description of the experimental design, data collection, and primary results. The current study focuses specifically on comparing different dimension reduction techniques for neural timeseries data.

Participants

Participants (N = 68) were business executives who attended "Summit LA," which is an organizational festival that invites successful founders, CEOs, and other executives to build community and discuss ideas for a weekend in Los Angeles. Of the 68 participants, 47 identified as male and 21 identified as female. One participant did not fill out the post-study survey necessary for analysis, leaving a final sample of N = 67 (mean age = 41.5 years old, SD = 9.87).

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Procedure

Participants who provided consent were invited into a pop-up neuroimaging lab in an office suite near the festival and fitted with an fNIRS cap. The experimental session consisted of two parts. First, participants engaged in a group interaction task where they took turns pitching each other ideas. The current study uses data from the second half of the session, unrelated to the group interaction task. In this phase of the experiment, participants' brains were scanned with fNIRS while they watched a video compilation of other executives talking about their attitudes toward work (Figure 1). After watching the videos, participants completed a post-study survey with a battery of questions probing their thoughts and feelings toward their career.

Outcome measures

After watching the videos, participants responded to four outcome measures that probed negative social-emotional sentiments toward their career: "I need a new or different challenge," "I feel overwhelmed," "I feel underappreciated," and "I feel burned out." Participants were instructed to respond to the measures as if they were thinking about their career in general. We used a 7-point Likert scale from "strongly disagree" to "strongly agree." We selected these specific outcomes because we believed them to be generally representative of aversive work-related experiences yet distinct enough from one another to capture different elements of the phenomenon.

To prepare data for machine learning, responses were binarized into *"*high*"* and *"*low*"* groups. Midpoints of response distributions were chosen to ensure near-balanced group distributions. The point at which the data was most evenly split for "I feel burned out" and "I

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

feel overwhelmed” was 3.5, which produced a ratio of 36 to 31 and 35 to 32 for their respective high and low groups. The other measures required that we split the data at integers to maintain relatively evenly balanced classes, so we had to omit the cases that fell on that midpoint. The midpoint for “I need a new or different challenge” was 3. After omitting the 5 participants that responded “3,” the sample size for this prediction model was 62 participants (32 high, 30 low). The final outcome measure, “I feel underappreciated,” was considerably skewed to the right. The most even split for this measure was at 2, which reduced the sample size for this model to 46 participants (27 high, 19 low). Given the relatively large reduction in sample size, imbalanced classes, and the difficulty of interpreting a split at 2 on a scale from 1-7, we decided to omit this measure (“I feel underappreciated”) from further analysis. Finally, minority groups were upsampled via randomized repetition to address class imbalance.

fNIRS data acquisition and preprocessing

We used NIRx NIRScout machines with a sampling rate of 3.906 Hz at wavelengths of 760 and 850 nm. Optodes on the fNIRS caps were arranged at 3-cm average source-detector separation distance over the prefrontal cortex (PFC), the temporal parietal junction (TPJ) and superior parietal lobule (Figure 2). 10-10 UI external positioning system was used to standardize the layout of our probes.

fNIRS data was sent through a preprocessing pipeline (Binnquist & Burns, 2021) that utilized custom scripts along with scripts from the Homer2 software (Huppert et al., 2009). For noisy and oversaturated channels, a modified quartile coefficient of dispersion (Bonett, 2006) was used to remove any channel that exceeded the coefficient cutoff ($C_{thresh} = 0.6 - 0.03 *$sampling rate) for two seconds or more. Further refinement of the data included

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

corrections for motion and non-neural changes in blood oxygenation. To address motion artifacts within remaining channels, spikes greater than 5 standard deviations change in less than 1 second were removed. To address non-neural physiological influences (e.g., cardiac and respiratory rhythms) and baseline drift, a conservative bandpass filter (0.008-0.2 Hz) was applied. Filtered data were then transformed from optical density to hemoglobin concentration values using the modified Beer Lambert Law with a standard differential path length filter. The final quality control step involved an autocorrelation change assessment to gauge the impact of motion correction. Channels displaying a substantial change in autocorrelation (exceeding a threshold of $r = 0.1$) were deemed significantly influenced by motion and excluded from subsequent analyses.

Several steps were then employed to reduce dimensionality and smooth the data during preprocessing. First, we selected specific brain ROIs. Given neural synchrony research implicating the default mode network in semantic and socio-cognitive processing (Lieberman, 2022), we focused our search to channels over the bilateral TPJ, the amPFC, and the dmPFC. Once the channels were averaged into their respective ROIs, we resampled the timeseries for each ROI from 3.906 Hz to one sample per second to further smooth our signal.

Model training and prediction

This research compared analytic approaches for neural timeseries-based machine learning models. For each model, our analysis process can be broken down into two parts – dimension reduction and classification. Analyses primarily differed in how they reduced dimensionality.

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Dimension reduction. We applied three approaches: MTPA with elastic net feature selection, MTPA with PCA feature extraction, and a down-sampling method using windowed averages. These approaches are also compared to a model which used the entire neural timeseries without dimension reduction. Because the scope of this paper was to explore and compare various dimension reduction approaches to enhance prediction with neural timeseries data, we applied each method separately. However, feature selection, feature extraction, and down sampling could be used in tandem to various degrees.

MTPA with elastic net feature selection. The feature selection-based MTPA was developed for Goldstein et al. (2025) to predict whether people feel overwhelmed, burned out, or in need of a new or different challenge with this same dataset. In this case, the temporal searchlight was performed with elastic net regression – a regularized regression technique that combines penalties from lasso (L1) and ridge (L2) to determine the best features. Lasso alone tends to delete features too easily if they are highly correlated, as in a timeseries (Zou & Hastie, 2005). On the other hand, ridge regression handles multicollinearity well, but it does not eliminate features. Elastic net combines the strengths of both methods – robust shrinkage and retention of useful correlated variables, making it particularly well-suited for high-dimensional datasets like those encountered in neuroimaging. By nature of being a feature selection approach, elastic net also allows us to map exact features back onto timepoints in the stimulus, offering the added benefit of exploring what is occurring in the stimulus when neural responses diverge between groups.

For these analyses, we used MATLAB's *lassoglm* to execute the elastic net feature selection. *lassoglm* determines the variables that are most representative of between group

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

differences by selecting the lambda, or shrinkage parameter representing the strength to which coefficients are penalized and reduced to 0, automatically through 10-fold cross-validation. Elastic net is also adjusted by a penalty parameter, alpha, which determines the relative contribution of the lasso and ridge penalties. For our purposes, the alpha parameter was manually set at 0.9 – closer to lasso to encourage as much sparsity as possible while still accounting for potential correlations among predictors. The combination of alpha and lambda parameters on the training set determined which features were retained. We then reduced the features of the training and testing timeseries to the selected features from the elastic net. All other features in the training and testing timeseries were discarded.

MTPA with PCA feature extraction. The feature extraction-based MTPA used principal component analysis (PCA) to reduce the dimensionality of the neural timeseries. PCA transforms high-dimensional data into a smaller set of orthogonal components while retaining as much variance as possible. The first component captures the direction of maximum variance in the data, the second component captures the maximum variance orthogonal to the first, and so on. PCA is advantageous because it creates a compact representation of the data, which can be particularly useful for high-dimensional datasets with limited sample sizes. However, PCA does not retain individual timepoints in the way feature selection methods do. While principal components can be back-projected to recover a rough correspondence with original features, this process is more indirect and less transparent than methods that retain the features in their original, immediately interpretable time-resolved form (Haufe et al., 2014).

We used MATLAB's *pca* function to perform PCA feature extraction. To ensure that we optimized the number of components while avoiding overfitting, we applied 10-fold cross-

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

validation within the training set to test the performance of models with different numbers of components. The number of components that achieved the highest cross-validated accuracy in the training data was then retained. Once the optimal components were identified, the training and testing timeseries were both projected onto this reduced set of components, and all other features were discarded.

Down sampling with windowed averages. For the down-sampling approach, we averaged over sets of features rather than selecting or transforming individual variables. By averaging adjacent timepoints within a window, this method helps to suppress noise and variability while retaining meaningful trends in the data. This can be especially useful for cognitive processes when fine-grained temporal fluctuations may not provide much predictive value, as may be the case for high-level cognitive and social processes.

To optimize the performance of this approach, we used 10-fold cross-validation within the training set to test a range of possible window sizes from 2 seconds to 1 minute. The window that yielded the best cross-validated accuracy was then applied to both the training and test datasets.

Classification via neural reference groups approach. The classification process was identical for all analyses compared here. We followed a neural reference groups approach (Dieffenbach et al., 2021) that trained models to recognize patterns of neural activity for people in two distinct groups and then classify new people based on the group with which they showed more neural similarity. Considering that the main goal of this paper was to evaluate how dimension reduction affects prediction with neural timeseries, we chose a logistic regression classifier over other models because of its linear framework and simplicity, though

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

different methods may also be employed with the same dimension reduction approaches. It is worth noting that while Dieffenbach et al. (2021) used Euclidean distance to classify between groups and we used logistic regression, the core concept of group assignment based on neural similarity remains the same.

We used leave-one-out cross validation to train and test our model. The leave-one-out process involved leaving out one subject at a time, performing dimension reduction and training the model on the remaining sample, and then testing the predictive performance of the model on the left-out subject. This was executed for every subject in our sample. To improve estimation stability, the entire process was then repeated 50 total times to account differences in overall prediction accuracy due to random shuffling during class balancing and fold assignment.

Significance testing

Permutation testing was used to assess how our actual predictive performance compared to what we would anticipate by chance. We conducted each approach described above 1000 times with shuffled labels. The outcome was a random chance-level distribution of cross-validated prediction accuracies representing the null hypothesis. Statistical significance was calculated by finding the number of permuted samples that are equal to or exceed the observed prediction accuracy, divided by the total number of iterations.

Experiment 1 Results and Discussion

Feature selection-based MTPA successfully predicted participants' psychological outcomes in four out of nine tests, spanning two of three measures (feeling overwhelmed and needing a new or different challenge) and all three ROIs (Figure 3). Using neural activity from

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

the TPJ, the MTPA accurately predicted whether participants felt overwhelmed with 72.84% accuracy ($p = 0.004$). Neural activity from the dmPFC also predicted whether people felt overwhelmed with 64.51% accuracy ($p = 0.034$). The strongest predictive performance overall occurred when neural activity in the dmPFC was used to predict whether participants felt in need of a new or different challenge in their careers (79.13%, $p < 0.0001$). The amPFC also showed predictive power approaching significance for this outcome, with accuracy reaching 63.29% ($p = 0.058$). These data were reported in Goldstein et al. (2025).

To evaluate how feature selection-based MTPA performed relative to other approaches, we tested the prediction accuracy of a feature extraction-based MTPA using PCA, down sampling with windowed averages, and a model using the full timeseries without dimensionality reduction. Across these approaches, PCA, windowed averaging, and the no dimension reduction model each showed moderate accuracy only for predicting whether people were burned out using the dmPFC. Prediction accuracies were 66.60% ($p = 0.011$), 63.43% ($p = 0.037$), and 65.79% ($p = 0.014$) for PCA, windowed averaging, and the full timeseries models without feature selection, respectively. These approaches did not predict any other outcomes or brain measures.

When directly comparing prediction accuracies across approaches, elastic net outperformed all other approaches for four out of five significant tests. PCA performed best on the remaining test, followed closely by no dimension reduction. Specifically, when predicting whether someone felt overwhelmed using the TPJ, elastic net achieved higher accuracy than PCA, windowed averaging, and no dimension reduction by 27.59%, 24.18%, and 22.15%, respectively. When predicting whether someone felt overwhelmed using the dmPFC, elastic net

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

again outperformed PCA, windowed averaging, and no dimension reduction by 15.52%, 15.29%, and 15.05%. Elastic net exhibited the highest performance when predicting whether someone needed a new or different challenge with the dmPFC, exceeding PCA, windowed averaging, and no dimension reduction by 28.07%, 26.07%, and 19.16%. Elastic net also demonstrated higher accuracy than PCA, windowed averaging, and no dimension reduction when predicting whether someone needed a new or different challenge with the amPFC by 7.00%, 5.97%, and 4.03%. Conversely, when predicting whether someone was burned out with the dmPFC, elastic net produced lower accuracy compared to PCA, windowed averaging, and no dimension reduction by 18.72%, 15.55%, and 17.91%.

Results from this experiment demonstrate the advantages of MTPA models using elastic net feature selection. While feature extraction, down-sampling, and using the full timeseries each successfully predicted one out of nine possible ROI-measure combinations, feature selection-based MTPA successfully predicted four combinations spanning two out of three outcome measures and all target ROIs. These findings confirmed our hypotheses, offering preliminary evidence for MTPA's use as a tool for predicting complex psychological states with neural timeseries data.

Experiment 2 Methods

Experiment 2 aimed to predict two distinct outcomes: whether participants were experiencing experimentally-induced cognitive load while watching a video and whether participants were given prior contextual information about the protagonist's behavior prior to the video. Although the type of outcome measures and participant population differ from

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Experiment 1, the model training, prediction, and significance testing remain identical to ensure consistency and support the generalizability of the previous findings.

Participants

Participants (N = 92) were undergraduate students recruited through the Psychology student subject pool. 20 identified as male, 71 as female, and 1 as non-binary. The mean age was 23.16 years (SD = 6.00). We aimed to recruit 30 participants per condition, with a slight oversampling to account for potential data loss. Due to equipment failure, data from one participant were excluded, resulting in 91 participants included in the final analysis. Participants were required to be fluent English speakers, have normal or corrected-to-normal vision, and be naïve to the movie clip used as a stimulus.

Procedure

All participants were scanned with fNIRS while they watched a three-minute scene from the movie “Brothers.” The video depicted a character behaving aggressively while other characters become upset and try to restrain him. The protagonist’s behavior could be interpreted through either situational or dispositional factors, and the video was displayed without sound to create more variability in interpretation. Participants viewed the clips under the following conditions: no-context, context, and context/dual-task (Figure 4). In the no-context condition, participants were simply instructed to watch the video closely. In the context condition, participants were told “This man has just learned that his wife has been cheating on him with his best friend.” Participants in the context/dual-task condition were given the same prompt, but they were also instructed to memorize a series of eight random digits prior to the

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

video, repeat these numbers in their head during the video, and recall the numbers in the correct order after the video. This simple dual-task design was used to induce cognitive load.

Outcome measures

We aimed to predict central aspects of participants' condition assignment – whether they were experiencing cognitive load and whether people received context for the protagonist's behavior. To ensure a balanced sample and reduce confounds, each outcome only utilized two out of the three conditions. When predicting whether people were experiencing cognitive load, we used the context and the context/dual-task conditions as our two classes. Both classes were consistent in that they received context, differing only in whether they experienced cognitive load. When predicting whether people received context, the context and no context conditions were used. In this case, neither class performed the dual-task, so they differed only in whether they received context.

fNIRS data acquisition and preprocessing

We used NIRx NIRSport2 machines with a sampling rate of 3.8147 Hz at wavelengths of 760 and 850 nm. The probe layout was comprised of 32 light sources and 32 detectors with a 3-cm average source-detector separation distance over the entire head (Figure 5). 10-10 UI external positioning system was used to standardize the layout of our probes.

fNIRS data was sent through the same preprocessing pipeline described in Experiment 1. Though we had full-head coverage with this montage, we selected the same ROIs as in the previous experiment to limit comparisons and remain consistent.

Experiment 2 Results and Discussion

Feature selection-based MTPA using elastic net regression successfully predicted both

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

outcomes – whether someone was experiencing cognitive load while they watched a video and whether someone was given context for a character's behavior prior to the video (Figure 6). First, neural activity in the amPFC predicted whether participants were experiencing cognitive load with 66.53% accuracy ($p = 0.026$). Models using neural activity from the TPJ predicted whether people had received contextual information about the protagonist prior to the video with 65.93% accuracy ($p = 0.024$).

Beyond the elastic net approach, models using feature extraction-based MTPA with PCA were able to predict whether participants received narrative context with neural activity from the amPFC (64.75%, $p = 0.031$). The windowed averaging approach produced the next highest prediction accuracy for the same ROI-measure combination (62.30%, $p = 0.049$). The full timeseries did not achieve significant predictive performance for any of the measures or ROIs.

When directly comparing prediction accuracies across approaches, elastic net outperformed all other approaches for two out of three significant tests, while PCA performed most effectively for the remaining test. Specifically, when predicting whether someone was experiencing cognitive load with the amPFC, elastic net achieved higher accuracy than PCA, windowed averaging, and no dimension reduction by 20.40%, 13.06%, and 15.23%, respectively. When using the TPJ to predict whether someone was given context prior to watching the video, elastic net also outperformed PCA, windowed averaging, and no dimension reduction by 9.69%, 6.72%, and 8.76%. However, when using the amPFC to predict whether someone was given context prior to watching the video, PCA outperformed the elastic net, windowed averaging, and no dimension reduction by 4.52%, 2.45%, and 3.27%.

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Overall, these findings replicate and extend the findings from Experiment 1. PCA, and to a lesser extent windowed averaging, was effective in predicting one measure with one ROI, but the elastic net approach was able to predict both outcome measures using two out of three ROIs. By successfully predicting both cognitive load and prior narrative context from neural data, MTPA proves to be an effective method for identifying predictive neural patterns in naturalistic settings above and beyond alternative methods or no temporal dimension reduction. These findings also support MTPA's general effectiveness across different types of measures and participant populations.

Additionally, by demonstrating that the amPFC predicts whether someone is experiencing cognitive load or not while the TPJ predicts narrative context, the elastic net MTPA revealed patterns that align with emerging theories differentiating sub-areas of the default mode network (Lieberman, 2022). Specifically, the CEEing model hypothesizes that TPJ will be more sensitive to narrative framing than cognitive load due to its role in integrating sensory and non-sensory information into pre-reflective construals, whereas mPFC regions will be more sensitive to cognitive load given their role in more effortful reflective processes. We observed patterns consistent with these predictions using MTPA with elastic net: TPJ yielded the highest prediction accuracy for narrative context (65.93%) but performed less effectively for cognitive load (58.07%), whereas amPFC achieved its highest accuracy for cognitive load (66.53%) and lower accuracy for narrative context (60.23%). These region-specific effects were only apparent when using the MTPA approach with elastic net, underscoring the advantage of applying feature selection-based dimension reduction techniques when working with neural timeseries data.

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Discussion

This study introduces and validates MTPA as an effective method for reducing the temporal dimensionality of neural timeseries data into a sparse and interpretable set of features for machine learning. Across two experiments, feature selection-based MTPA using elastic net regression consistently outperformed alternative and more conventional methods, predicting six out of twelve possible tests. PCA and windowed averaging performed effectively on two occasions. Using the full timeseries without feature selection was successful only once. These findings demonstrate the broad efficacy and applicability of using MTPA prior to classification, specifically emphasizing the advantage of feature selection-based MTPA.

Approaches were compared across significant ROI-measure combinations, and elastic net outperformed all other methods in six out of eight significant tests, often with large differences (e.g., elastic net exceeded PCA, windowed averaging, and baseline by 27.6%, 24.2%, and 22.2%, respectively, when predicting whether someone needed a new or different challenge with the dmPFC). In contrast, PCA emerged as the top-performing method in two tests, suggesting that PCA can serve as a viable alternative in some cases although it mostly fell short of elastic net overall. Windowed averaging and no dimension reduction lagged elastic net or PCA in all comparisons.

Feature selection-based MTPA generalized well across both experiments despite their differences in methods and objective. Experiment 1 was conducted in the field with a real-world population of business professionals and successfully predicted persistent, oft-hidden social-emotional states such as feeling overwhelmed and needing a new or different challenge in one's career. Experiment 2 was conducted in a controlled laboratory setting with an

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

undergraduate participant population and focused on cognitive load and narrative context during video viewing. MTPA with elastic net successfully predicted outcome measures in all three tests. Feature selection-based MTPA's ability to generalize across distinct populations (real-world and lab-based) and outcome types (persistent psychological states, cognitive processes, and contextual understanding) highlights its strength and versatility.

Interpretability and practical applications

Beyond its higher predictive power compared to other methods explored here, MTPA offers a second critical advantage – interpretability. By selecting specific timepoints that contribute to group-level differences, elastic net allows researchers to directly map neural responses back to precise moments in a stimulus. This temporal specificity allows us to reverse-engineer the specific timepoints in the video stimulus that evoke distinct neural profiles between groups, offering insights into the connection between stimulus content, neural activity, and psychological lenses (for an example, see Goldstein et al., 2025). In contrast, PCA transforms the data into components that maximize variance but lose direct connection to the original timepoints. While PCA achieved moderate prediction accuracy in some cases, its lower temporal interpretability compared to feature selection limits its utility for understanding the dynamics of neural responses during complex, naturalistic stimuli. Similarly, methods like windowed averaging smooth over meaningful temporal fluctuations and fail to leverage the richness of long neural timeseries. We believe this is another reason why MTPA with elastic net feature selection is the optimal choice for most cases using naturalistic stimuli.

Conclusion

The ability to analyze neural responses from extended naturalistic designs has become a

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

central goal in social neuroscience. However, while these stimuli afford ecologically valid experiences that closely approximate real-world processes, their length and resulting temporal complexity pose major computational challenges due to overfitting and poor generalization. This study demonstrates that MTPA alleviates these challenges by reducing the timeseries into sparse yet interpretable patterns of timepoints that improve predictive power. Consistent findings from the two experiments showed that feature selection-based MTPA predicts a variety of measures above and beyond PCA, windowed averaging, and no temporal dimension reduction – approaches that all either reduce psychological interpretability, aggregate over important variability, or use overwhelmingly long neural timeseries. As naturalistic stimuli become increasingly prevalent, methods like MTPA will be essential for leveraging the brain to predict how people think, feel, and behave in dynamic, real-world contexts.

Acknowledgements

We thank the Summit Series, L.L.C. for granting us access to the festival space and attendees for Experiment 1. We would also like to thank Richard Barbour, Lamija Pašalić, and Angelo Herrera from NIRx for various forms of technical support. We thank Dan Ames, Razia Sahi, Kevin Tan, Grace Gillespie, Eugene Hsu, Patrycja Szafran, and Maisy Tarlow for all their hard work in collecting the data in a short period of time. For their help with data collection in Experiment 2, we would like to thank undergraduate research assistants Malia Groth, Tracy Mao, Laura Li, Howard Fung, and Brandon Ha. Additionally, we thank Dr. Hongjing Lu of the UCLA Computational Vision and Learning Lab for her coursework and guidance that inspired the initial ideas for this project.

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Funding and ethics

The entire project was made possible by funding from the National Defense Science and Engineering Graduate (NDSEG) Fellowship. Experiment 1 was supported by an anonymous donor. Experiment 2 was supported by unrestricted funds. This experiment received ethical approval by UCLA Institutional Review Board (#22-001785).

Disclosure of interest

The authors declare no competing interests.

Data availability

Data for these experiments can be accessed online at https://osf.io/b84c3/.

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

References

Binnquist, A., & Burns, S. (2021). *SCN fNIRS preprocessing pipeline* (Version 2.0.0) [MATLAB]. https://github.com/abinnquist/fNIRSpreProcessing/tree/main (Original work published 2021)

Bonett, D. G. (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis*, *50*(11), 2953–2957. https://doi.org/10.1016/j.csda.2005.05.007

Burns, S. M., & Lieberman, M. (2019). *The Use of fNIRS for Unique Contributions to Social and Affective Neuroscience*. OSF. https://doi.org/10.31234/osf.io/kygbm

Concato, J., Peduzzi, P., Holford, T. R., & Feinstein, A. R. (1995). Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *Journal of Clinical Epidemiology*, *48*(12), 1495–1501. https://doi.org/10.1016/0895-4356(95)00510-2

Dieffenbach, M. C., Gillespie, G. S. R., Burns, S. M., McCulloh, I. A., Ames, D. L., Dagher, M. M., Falk, E. B., & Lieberman, M. D. (2021). Neural reference groups: A synchrony-based classification approach for predicting attitudes using fNIRS. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 117–128. https://doi.org/10.1093/scan/nsaa115

Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215*, 116828. https://doi.org/10.1016/j.neuroimage.2020.116828

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Goldstein, B. M., Burns, S. M., Dieffenbach, M. C., Konkoly, C., Abramowitz, S., & Lieberman, M. D. (2025). *Tell me you're overwhelmed at work without telling me you're overwhelmed: Detecting the neural predictors of hidden and persistent psychological states*. PsyArXiv. https://doi.org/10.31234/osf.io/u7g9a_v1

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(7–8), 1157–1182. https://doi.org/10.1162/153244303322753616

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Hua, J., Tembe, W. D., & Dougherty, E. R. (2009). Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognition*, *42*(3), 409–424. https://doi.org/10.1016/j.patcog.2008.08.001

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*(10), D280-298. https://doi.org/10.1364/ao.48.00d280

Jääskeläinen, I. P., Sams, M., Glerean, E., & Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage*, *224*, 117445. https://doi.org/10.1016/j.neuroimage.2020.117445

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical*

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

and Engineering Sciences, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X

Li, L., Huang, X., Xiao, J., Zheng, Q., Shan, X., He, C., Liao, W., Chen, H., Menon, V., & Duan, X. (2022). Neural synchronization predicts marital satisfaction. *Proceedings of the National Academy of Sciences*, *119*(34), e2202515119. https://doi.org/10.1073/pnas.2202515119

Lieberman, M. D. (2022). Seeing minds, matter, and meaning: The CEEing model of pre-reflective subjective construal. *Psychological Review*, *129*(4), 830–872. https://doi.org/10.1037/rev0000362

Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience and Biobehavioral Reviews*, *99*, 311–328. https://doi.org/10.1016/j.neubiorev.2018.12.021

Mwangi, B., Tian, T. S., & Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, *12*(2), 229–244. https://doi.org/10.1007/s12021-013-9204-3

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, *9*(1), 332. https://doi.org/10.1038/s41467-017-02722-7

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373–1379. https://doi.org/10.1016/s0895-4356(96)00236-3

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*(3), 147–156. https://doi.org/10.1016/j.tics.2012.01.005

Scholkmann, F., Kleiser, S., Metz, A. J., 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. https://doi.org/10.1016/j.neuroimage.2013.05.004

Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology*, *165*(6), 710–718. https://doi.org/10.1093/aje/kwk052

Weaverdyck, M. E., Lieberman, M. D., & Parkinson, C. (2020). Tools of the Trade Multivoxel pattern analysis in fMRI: A practical introduction for social and affective neuroscientists. *Social Cognitive and Affective Neuroscience*, *15*(4), 487–509. https://doi.org/10.1093/scan/nsaa057

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same Story, Different Story. *Psychological Science*, *28*(3), 307–319. https://doi.org/10.1177/0956797616682029

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *67*(2), 301–320.

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Captions to figures

Figure 1. Stills from the experiment and video stimulus. The left image shows participants wearing fNIRS and watching the video together. The right image is a still from the video stimulus, where other business executives were instructed to speak about the positive and negative aspects of their work. Captions were not included in the original stimulus.

Figure 2. Montage of fNIRS probes across the prefrontal cortex, temporal parietal junction, and superior parietal lobule. The red dots represent source optodes and the blue dots represent detector optodes. ROIs are highlighted in orange (amPFC), purple (dmPFC), and green (TPJ).

Figure 3. Heatmap of prediction accuracies for each analysis approach, ROI, and outcome measure. Note: ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05, ~p<.0.06

Figure 4. Instructions provided to participants in each of the three conditions for Experiment 2.

Figure 5. Montage of fNIRS probes for the 108-channel full-head setup. The red dots represent source optodes and the blue dots represent detector optodes. Red dots encircled with blue rings represent short-channel detectors which aid in preprocessing. ROIs are highlighted in orange (amPFC), purple (dmPFC), and green (TPJ).

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

Figure 6. Heatmap of prediction accuracies for each analysis approach, ROI, and outcome measure. Note: *p<0.05

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

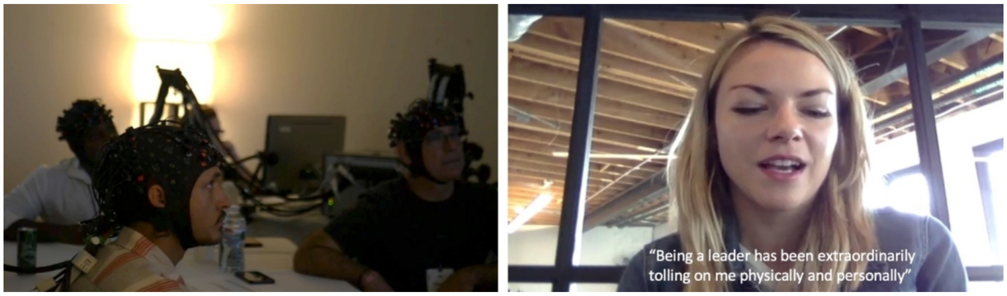

Figure 1. Stills from the experiment and video stimulus. The left image shows participants wearing fNIRS and watching the video together. The right image is a still from the video stimulus, where other business executives were instructed to speak about the positive and negative aspects of their work. Captions were not included in the original stimulus.

140x40mm (300 x 300 DPI)

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

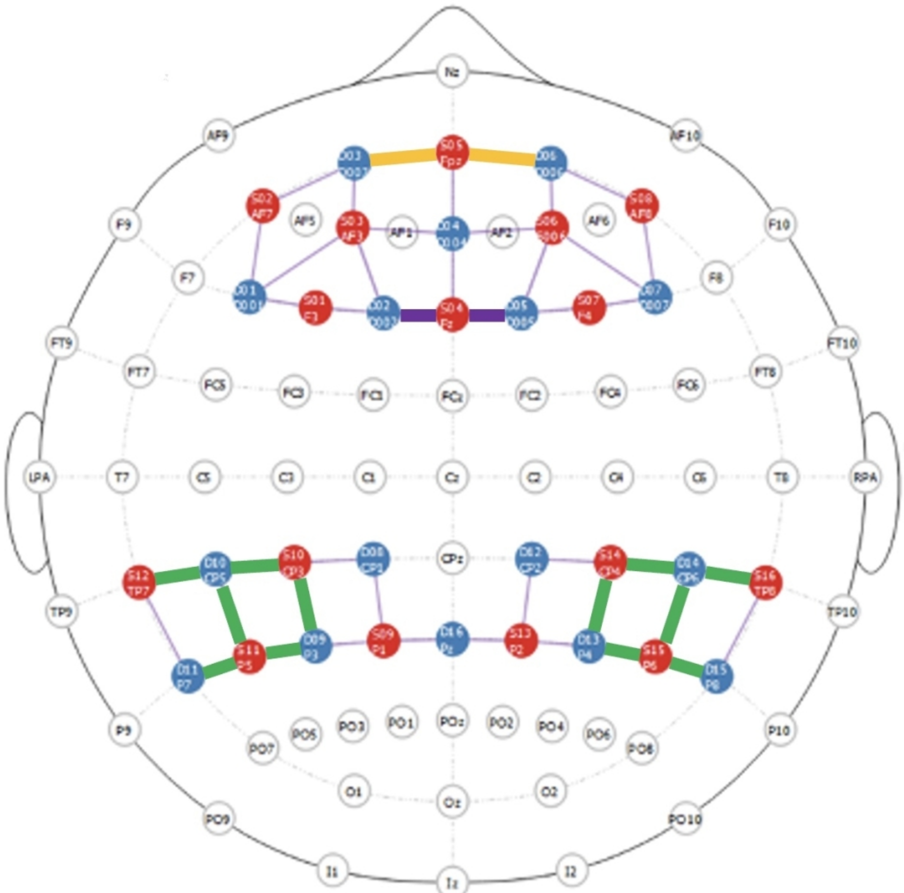

Figure 2. Montage of fNIRS probes across the prefrontal cortex, temporal parietal junction, and superior parietal lobule. The red dots represent source optodes and the blue dots represent detector optodes. ROIs are highlighted in orange (amPFC), purple (dmPFC), and green (TPJ).

113x106mm (300 x 300 DPI)

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

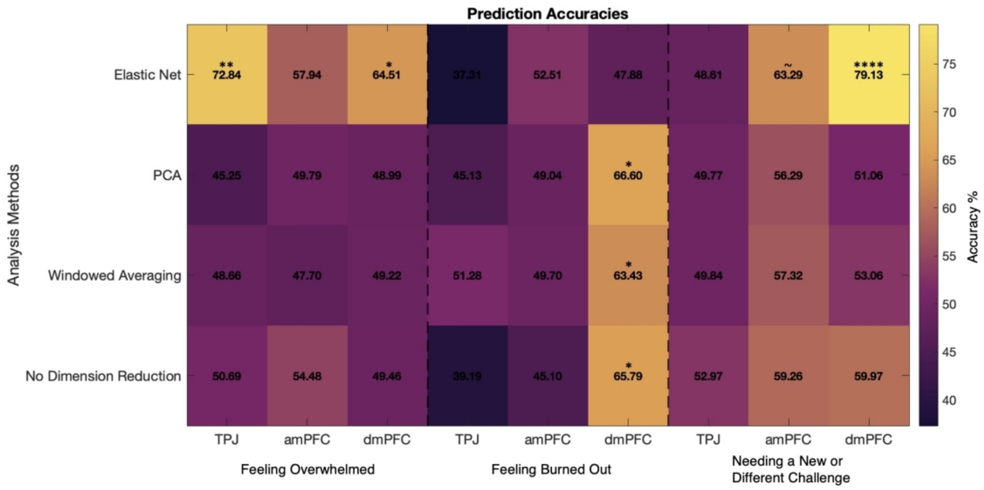

Figure 3. Heatmap of prediction accuracies for each analysis approach, ROI, and outcome measure. Note: ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05, ~p<.0.06

161x83mm (300 x 300 DPI)

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

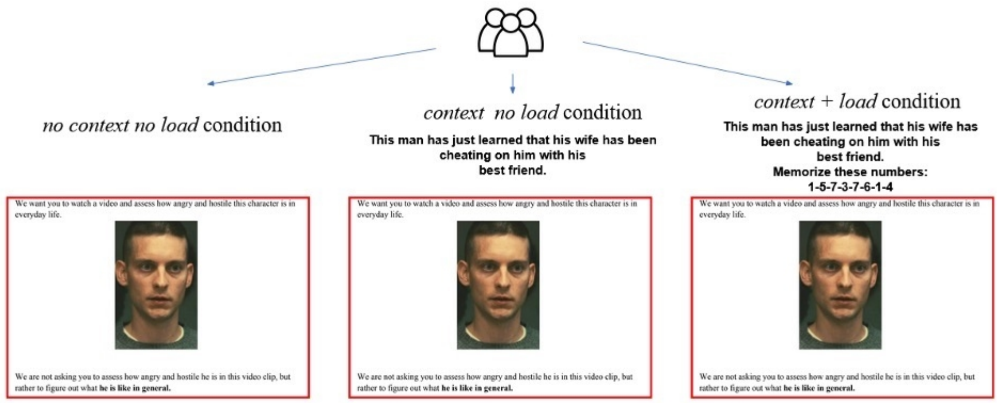

Figure 4. Instructions provided to participants in each of the three conditions for Experiment 2.

79x36mm (300 x 300 DPI)

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

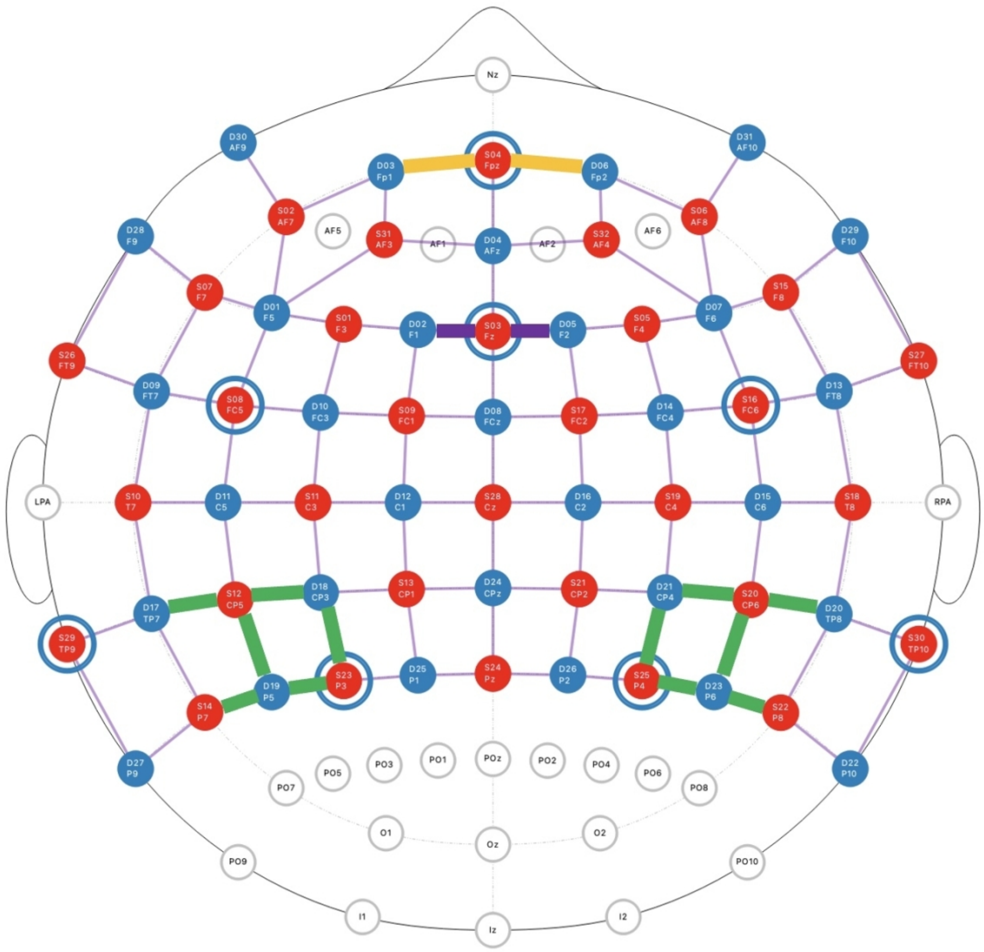

Figure 5. Montage of fNIRS probes for the 108-channel full-head setup. The red dots represent source optodes and the blue dots represent detector optodes. Red dots encircled with blue rings represent short-channel detectors which aid in preprocessing. ROIs are highlighted in orange (amPFC), purple (dmPFC), and green (TPJ).

109x107mm (300 x 300 DPI)

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025

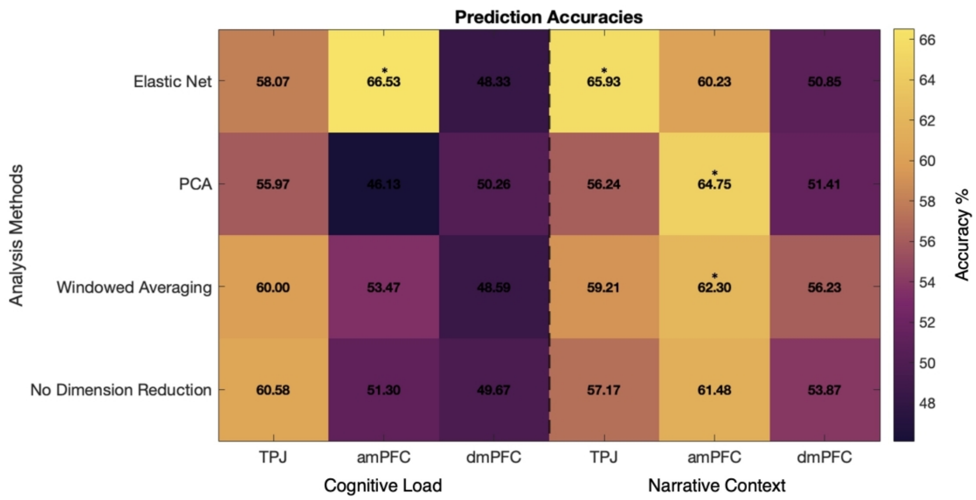

Figure 6. Heatmap of prediction accuracies for each analysis approach, ROI, and outcome measure. Note: *p<0.05

199x102mm (300 x 300 DPI)

Downloaded from https://academic.oup.com/scan/advance-article/doi/10.1093/scan/nsaf058/8160481 by guest on 07 July 2025