TRANSLATING NEUROPHYSIOLOGICAL RECORDINGS INTO DYNAMIC ESTIMATES OF CONCEPTUAL KNOWLEDGE AND LEARNING

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Bachelor of Arts

in

Mathematics

by

Daniel Carstensen

Advised by Jeremy Manning and Peter Mucha

Dartmouth College

Hanover, New Hampshire

May 24th 2024

Abstract

This study investigates the potential of a computational approach to provide moment-by-moment insights into a student's comprehension of lecture material through analysis of their neurophysiological responses during the lecture. In doing so, we present a solution to two difficult problems. How do we quantify the conceptual content of a lecture video? And how can we use EEG recordings to compute a knowledge estimate of this conceptual content? First, we used topic modeling to generate moment-by-moment estimates of the conceptual content of a lecture. Then we used EEG recordings collected during the lecture to compute an ISFC-derived knowledge estimate of this conceptual content. We found that particularly gamma band activity may contain a signal indicative of knowledge acquisition.

Acknowledgments

First and foremost, I would like to extend my gratitude to all the members of the Context Lab, in particular to Dr. Jeremy Manning and Paxton Fitzpatrick who have continuously supported my research journey throughout my time at Dartmouth. I would also like to thank Dr. Peter Mucha for his valuable advice and support throughout this project. I couldn't have completed this thesis without them.

I also want to thank my friends and family members who have supported me both academically and emotionally.

Contents

Chapter 1

Introduction

Section 1.1

Feedback: a Critical Component of Instruction

For most of modern history, education has followed a similar format. An instructor and a number of students gather in the same place. There, the instructor lectures, answers questions, and, crucially, provides feedback on each student's progress. What has the student understood well and which parts of the lecture should they review? Naturally, such continuous feedback is critical to the overall learning progress of each student and has traditionally taken the form of a quiz that tests the students' knowledge of a certain number of concepts. However, creating high-quality quiz questions is a time-consuming effort. Even once a quiz is completed, the time needed to provide fine-scale feedback to each student limits the overall number of students with which an instructor can work.

In recent years, the explosive growth of widely available pre-recorded lectures on platforms such as YouTube has highlighted this fundamental educational bottleneck. Although anyone can now find high-quality lectures on almost any topic online, such lectures rarely come with a set of quiz questions that tests conceptual understanding

of the presented concepts.

We took a step towards a possible solution to this problem that could allow us to provide precise feedback to a student after they watched a lecture while eliminating the need for testing. In particular, we developed a computational approach that, after further development and testing, could provide instantaneous feedback on a student's conceptual understanding of a lecture based on the student's neurophysiological response to watching the lecture.

Section 1.2

Background

Our approach was heavily informed by previous work by Fitzpatrick et al. [2023] who have developed a computational framework to estimate the conceptual knowledge of learners during training by using multiple-choice quizzes. Specifically, they parsed the transcript of a lecture video into overlapping sliding windows and mapped the text snippet in each window to a high-dimensional topic vector. They showed that the resulting sequence of topic vectors yields the lecture's conceptual trajectory over time within the topic space. They then mapped the questions into the topic space which allowed them to estimate the relevance of any timepoint in the lecture to a given question. This result was crucial to the present work as it provided us with a tool to estimate the conceptual trajectory of a lecture and identify those moments which are crucial to the understanding of a given question.

To translate neurophysiological recordings into knowledge and learning estimates, we further needed a computational method that has shown promising results when applied to a naturalistic learning task. Prior studies by Hasson et al. [2004] and Simony et al. [2016] have used inter-subject correlations (ISC) and inter-subject functional correlations (ISFC) between time-aligned neurophysiological signals recorded

from different individuals to differentiate stimulus-driven dynamics from non-stimulus driven dynamics. Crucially, Poulsen et al. [2017] extended the use of ISC to a classroom setting in which they collected the EEG signal of multiple participants who were watching a video clip. They showed that, under the assumption that ISC is a proxy for attention, stimulus-evoked neural responses can be tracked through EEG data.

Section 1.3

Mapping Brain Activity to Lecture Concepts

Our main objective was to develop a computational approach that could provide moment-by-moment knowledge estimates of the content presented in a lecture based on neurophysiological data recorded while a participant was watching the lecture. Hence, we needed to solve two problems. First, how do we quantify the conceptual content that is presented at any given point in a lecture video? Second, how can we use raw EEG recordings to compute a knowledge estimate of this conceptual content?

We solved the first problem following the approach of Fitzpatrick et al. [2023]. Using a topic model, we generated moment-by-moment text embeddings of a lecture's transcript in the form of topic vectors. These topic vectors served as a quantified estimate of the conceptual content of any moment in a lecture (see Section 2.2.1). We solved the second problem by extrapolating earlier work by Hasson et al. [2004], Poulsen et al. [2017], and Simony et al. [2016]. Instead of using ISFCs as a proxy for attention, we treated them as estimates of learning. In combination, this allowed us to use a participant's ISFC corresponding to a moment in a lecture as an estimate of the participant's conceptual understanding of the content presented at that moment (see Section 2.4).

Section 1.4

Notation and Terminology

We formally define the following terms. To gain a conceptual understanding of each term and how it relates to the experiment and analysis, see Sections 2.1.1, 2.1.2, 2.2.1, and 2.3.

- We will use two shorthand notations when referring to the two Khan Academy lecture videos we selected. We will refer to the lecture *Earth Formation* as Lecture 1 in text and as EF in mathematical writing. We will refer to the lecture *Plate Tectonics: Difference between crust and lithosphere* as Lecture 2 in text and as PT in mathematical writing.

- We represent the set of all quiz questions as the set

$$Q = \{\text{Q1, Q2, } \cdots \text{, Q90}\}.$$

  Further, we define the set of Lecture 1 quiz questions, the set of Lecture 2 quiz questions, and the set of General Knowledge quiz questions as the subsets

$$\begin{aligned} Q_{\text{EF}} &= \{\text{Q}k : 1 \leq k \leq 30\} \subset Q, \\ Q_{\text{PT}} &= \{\text{Q}k : 31 \leq k \leq 60\} \subset Q, \\ Q_{\text{General}} &= \{\text{Q}k : 61 \leq k \leq 90\} \subset Q, \end{aligned}$$

  respectively.

- We represent the set of all participants by the set

$$P = \{\text{P1, P2, } \cdots \text{, P42}\}.$$

- We represent the *topic vector* of some document L with k topics as the k-dimensional row vector

$$\mathbf{tv}^{\mathrm{L}} = \begin{pmatrix} w_1^{\mathrm{L}} & w_2^{\mathrm{L}} & \cdots & w_k^{\mathrm{L}} \end{pmatrix} \in \mathcal{M}_{1\times k}(\mathbb{R}).$$

- We represent the *trajectory* of a lecture L of length t embedded in a k-dimensional topic space as the $t \times k$ matrix

$$\mathbf{T}_{\mathrm{L}} = \begin{pmatrix} \mathbf{tv}_1^{\mathrm{L}} \\ \mathbf{tv}_2^{\mathrm{L}} \\ \vdots \\ \mathbf{tv}_t^{\mathrm{L}} \end{pmatrix} = \begin{pmatrix} w_{1,1}^{\mathrm{L}} & w_{1,2}^{\mathrm{L}} & \cdots & w_{1,k}^{\mathrm{L}} \\ w_{2,1}^{\mathrm{L}} & w_{2,2}^{\mathrm{L}} & \cdots & w_{2,k}^{\mathrm{L}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{t,1}^{\mathrm{L}} & w_{t,2}^{\mathrm{L}} & \cdots & w_{t,k}^{\mathrm{L}} \end{pmatrix} \in \mathcal{M}_{t\times k}(\mathbb{R}),$$

  where $\mathbf{tv}_t^L$ is the 22-dimensional topic vector at time t in the corresponding lecture L.

- We represent the broadband power of a *single-channel EEG recording* of participant P$i \in P$ recorded while they were watching lecture L of length t as the t-dimensional column vector

$$\mathbf{c}^{\mathrm{P}i,\mathrm{L}} = \begin{pmatrix} c_1^{\mathrm{P}i,\mathrm{L}} & c_2^{\mathrm{P}i,\mathrm{L}} & \cdots & c_t^{\mathrm{P}i,\mathrm{L}} \end{pmatrix}^{\mathrm{T}}.$$

  To represent the *narrowband power* of $\mathbf{c}^{\mathrm{P}i,\mathrm{L}}$ for a predefined frequency band η, we write

$$\mathbf{c}_{\boldsymbol{\eta}}{}^{\mathrm{P}i,\mathrm{L}} = \begin{pmatrix} c_{1;\eta}^{\mathrm{P}i,\mathrm{L}} & c_{2;\eta}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{t;\eta}^{\mathrm{P}i,\mathrm{L}} \end{pmatrix}^{\mathrm{T}} \in \mathcal{M}_{k\times 1}(\mathbb{R}).$$

- We represent the broadband power of a *k-channel EEG recording* of participant P$i \in P$ recorded while they were watching lecture L of length t as the $t \times k$

matrix

$$\mathbf{E}^{\mathrm{L}}_{\mathrm{P}i} = \begin{pmatrix} (\mathbf{c}_1^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} & (\mathbf{c}_2^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} & \cdots & (\mathbf{c}_k^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} \end{pmatrix}$$
$$= \begin{pmatrix} c_{1,1}^{\mathrm{P}i,\mathrm{L}} & c_{1,2}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{1,k}^{\mathrm{P}i,\mathrm{L}} \\ c_{2,1}^{\mathrm{P}i,\mathrm{L}} & c_{2,2}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{2,k}^{\mathrm{P}i,\mathrm{L}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{t,1}^{\mathrm{P}i,\mathrm{L}} & c_{t,2}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{t,k}^{\mathrm{P}i,\mathrm{L}} \end{pmatrix} \in \mathcal{M}_{t\times k}(\mathbb{R}).$$

To represent the "concatenation" of the broadband power signal of $\mathbf{E}^{\mathrm{L}}_{\mathrm{P}i}$ and m narrowband power signals extracted from each EEG channel in correspondence to a set of frequency bands $H = \{\eta_1, \eta_2, \cdots, \eta_m\}$, we define the block matrix

$$\mathbf{E}^{\mathrm{L}}_{\mathrm{P}i,\mathbf{H}} = \left(\begin{array}{cccc|c|cccc} (\mathbf{c}_1^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} & (\mathbf{c}_{1,\boldsymbol{\eta}_1}^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} & \cdots & (\mathbf{c}_{1,\boldsymbol{\eta}_m}^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} & \cdots & (\mathbf{c}_k^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} & (\mathbf{c}_{k,\boldsymbol{\eta}_1}^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} & \cdots & (\mathbf{c}_{k,\boldsymbol{\eta}_m}^{\mathrm{P}i,\mathrm{L}})^{\mathrm{T}} \end{array}\right)$$
$$= \left(\begin{array}{cccc|c|cccc} c_{1,1}^{\mathrm{P}i,\mathrm{L}} & c_{1,2;\boldsymbol{\eta}_1}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{1,m;\boldsymbol{\eta}_m}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{1,k(m-1)}^{\mathrm{P}i,\mathrm{L}} & c_{1,k(m-1)+1;\boldsymbol{\eta}_1}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{1,km;\boldsymbol{\eta}_m}^{\mathrm{P}i,\mathrm{L}} \\ c_{2,1}^{\mathrm{P}i,\mathrm{L}} & c_{2,2;\boldsymbol{\eta}_1}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{2,m;\boldsymbol{\eta}_m}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{2,k(m-1)}^{\mathrm{P}i,\mathrm{L}} & c_{2,k(m-1)+1;\boldsymbol{\eta}_1}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{2,km;\boldsymbol{\eta}_m}^{\mathrm{P}i,\mathrm{L}} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ c_{t,1}^{\mathrm{P}i,\mathrm{L}} & c_{t,2;\boldsymbol{\eta}_1}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{t,m;\boldsymbol{\eta}_m}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{t,k(m-1)}^{\mathrm{P}i,\mathrm{L}} & c_{t,k(m-1)+1;\boldsymbol{\eta}_1}^{\mathrm{P}i,\mathrm{L}} & \cdots & c_{t,km;\boldsymbol{\eta}_m}^{\mathrm{P}i,\mathrm{L}} \end{array}\right)$$

Note that $\mathbf{E}^{\mathrm{L}}_{\mathrm{P}i,\mathbf{H}} \in \mathcal{M}_{t\times(k\cdot m)}(\mathbb{R})$. We call both $\mathbf{E}^{\mathrm{L}}_{\mathrm{P}i}$ and $\mathbf{E}^{\mathrm{L}}_{\mathrm{P}i,\mathbf{H}}$ neural feature matrices.

Chapter 2

Methods

Section 2.1

Experimental Details

2.1.1. Experimental Protocol

We hand-selected two lecture videos from the Khan Academy course "Cosmology and Astronomy," namely *Earth Formation* (Lecture 1, duration: 10:08 min) and *Plate Tectonics: Difference between crust and lithosphere* (Lecture 2, duration: 7:59 min). Lecture 1 provides a brief overview of the early geological history of Earth. Lecture 2 explains the difference between the chemical and mechanical properties of the different layers of Earth. We selected these particular lectures to satisfy three general criteria as proposed by Fitzpatrick et al. [2023]. First, we wanted to make both lectures accessible to a wide audience, minimizing the need for prior knowledge. We achieved this by selecting lectures designed for students at the outset of their training in specific content areas. Second, we wanted both lectures to have related content, as we wanted to test the ability of our approach to differentiate similar conceptual content. To do so, we selected two videos from the "Cosmology and Astronomy" course domain on Khan Academy. Third, we intentionally avoided overlap between

the videos to prevent understanding of one from directly influencing comprehension of the other. To meet this criterion, we chose videos from distinct lecture series: Lecture 1 from the "Life on Earth and in the Universe" series and Lecture 2 from the "Earth Geological and Climatic History" series.

We also hand-wrote three sets of 30 multiple-choice quiz questions that we hoped would allow us to evaluate participants' knowledge of each individual lecture, along with related knowledge about geology not specifically presented in either video (see Appendix B, **??**). Each question set targets a particular content area, that is, the Lecture 1 question set tests understanding of the Earth Formation lecture, the Lecture 2 question set tests understanding of the Plate Tectonics lecture, and the General Knowledge question set tests knowledge of general concepts from geology. During each experiment session, we created three quizzes by drawing ten questions from each question set at random without replacement, so that each quiz contained 30 questions in total and each question appeared exactly once. We also randomized the orders of questions on each quiz and the orders of answer options for each question. Participants completed Quiz 1 before watching Lecture 1, Quiz 2 after watching Lecture 1 but before watching Lecture 2, and Quiz 3 after watching Lecture 2. Quiz 1 was intended to assess participants' "baseline" knowledge before training, Quiz 2 assessed knowledge after watching Lecture 1, and Quiz 3 assessed knowledge after watching Lecture 2.

Our experimental protocol was approved by the Committee for the Protection of Human Subjects at Dartmouth College.

2.1.2. EEG Recordings

We used a custom BrainVision actiCAP snap active electrode system coupled with an actiCHamp amplifier to record 64-channel EEG data in addition to single-channel HEOG (horizontal electro-oculogram) data at 500 Hz while participants complete the

experiment. We collected data from 46 participants, of which we discarded four. In particular, we discarded one data set since the participant reported that they had fallen asleep during the experiment. In addition, we discarded another data set since the participant did not follow the experiment instructions ($<$ 1 s response time on all questions). Finally, we discarded two data sets due to faulty timestamping which inhibits the correct alignment of the EEG data with the experiment time line.

Section 2.2

Joint Text Embeddings of Lectures and Questions

2.2.1. Topic Modeling

We followed the approach taken by Fitzpatrick et al. [2023] to create a joint text embedding space from both lectures and questions. In particular, we used Latent Dirichlet Allocation (LDA) (Blei et al. [2003]). In short, LDA is a generative probabilistic technique for uncovering thematic structure in document collections. It assumes that documents are mixtures of a predefined number of latent topics, each characterized by a probability distribution over a shared vocabulary. These topics themselves are probability distributions over words, allowing for vocabulary reuse across topics. The model employs a Bayesian framework, where documents are generated by first selecting a topic distribution and then drawing tokens according to the chosen topic's word distribution. By analyzing word co-occurrence patterns, LDA estimates the topic distributions for each document and each topic, revealing the underlying thematic landscape of the corpus.

Once fitted to a set of documents with k topics, LDA can generate a topic embedding for any document in the training set, or any new document that contains at least some of the words in the model's vocabulary. This embedding is in the form of a

k-dimensional vector that describes how much the document (most probably) reflects each topic. For a more detailed description of LDA see Appendix A, 5.1.1.

2.2.2. Text Preprocessing

To generate a collection of documents that we could use to fit an LDA model, we followed the approach taken by Fitzpatrick et al. [2023]. We retrieved the transcript for each lecture consisting of one timestamped line of text for every few seconds (mean: 2.98 s; standard deviation: 1.29 s) of spoken content in the video (i.e., corresponding to each individual caption that would appear on screen while watching the lecture, and when those lines would appear). We then defined a sliding window length of (up to) $w = 10$ transcript lines and assigned each window a timestamp corresponding to the endpoint of its last line's timestamp. These sliding windows ramped up and down in length at the very beginning and end of the transcript, respectively. In other words, the first sliding window covered only the first line of the transcript; the second sliding window covered the first two lines; and so on. This ensured that each line of the transcript appeared in the same number (w) of sliding windows.

After performing various standard text preprocessing (e.g., normalizing case, lemmatizing, removing punctuation and stop words) relying on several functionalities of the NLTK Python package (Bird et al. [2009]), we treated the text from each sliding window as a single document and combined these documents across the windows generated from each lecture to create a single training corpus for the topic model.

Overall, this approach yielded 192 windows for Lecture 1 and 181 windows for Lecture 2 for a total training corpus of 373 documents. To generate a shared vocabulary of all unique tokens in the lecture transcripts, we identified all unique words after preprocessing. In addition, we used the `Phrases` class provided by the Gensim Python package (Řehůřek and Sojka [2010]) to identify common bigrams. We chose to include bigrams to help LDA capture more specific topics by representing focused

concepts that single words alone might miss (e.g., "pacific_plate").

2.2.3. Fitting the Topic Model

Since the number of topics that LDA identifies is a tunable hyperparameter, we chose to employ two heuristics to find a suitable number of topics when training a Gensim `LdaModel`. Our first heuristic aimed to find the maximum number of topics such that each topic had a well-separated content. To achieve this, we evaluated a validation set of documents (excluded during training) and examined whether each topic had a within-topic variance of above 1e-4. A variance at or below this threshold indicates that a topic does not discriminate well between documents. Our second heuristic aimed to find the number of topics that corresponds to the highest CV coherence score evaluated by Gensim `CoherenceModel` on a validation set of documents (excluded during training). The coherence score measures the similarity of the top 20 words of each topic. Hence, a higher coherence score implies topics that are more interpretable by humans.

We used 5-fold cross-validation and took the floored average of the optimal number of topics found by each heuristic (23 and 22, respectively), yielding an overall number of 22 topics. The top 10 words from each of the 22 topics discovered can be found in Appendix B 6.1.

2.2.4. Generating Text Embeddings

Once we had fitted an LDA model, we could generate a 22-dimensional topic vector for an arbitrary (potentially unseen) document. An advantageous feature of these topic vectors is that documents sharing comparable topic mixtures (that is, documents that exhibit similar themes according to the model) will produce similar coordinates (using measures such as Euclidean distance, correlation, or other spatial metrics). Generally, comparing the topic vectors from various documents helps in identifying

their similarity in conceptual content.

We then transformed the text of each sliding window into a topic vector and used linear interpolation (independently for each topic dimension) to upsample the resulting time series to 10 Hz and match the length of its corresponding lecture video using the method `interp1d` from the Python package SciPy (Virtanen et al. [2020]). We also used the fitted model to obtain topic vectors for each question in our pool. Together, we obtained the trajectory matrices

$$\mathbf{T}_{\mathrm{EF}} = \begin{pmatrix} \mathbf{tv}_1^{\mathrm{EF}} & \mathbf{tv}_2^{\mathrm{EF}} & \cdots & \mathbf{tv}_{6283}^{\mathrm{EF}} \end{pmatrix}^{\mathrm{T}}$$
$$\mathbf{T}_{\mathrm{PT}} = \begin{pmatrix} \mathbf{tv}_1^{\mathrm{PT}} & \mathbf{tv}_2^{\mathrm{PT}} & \cdots & \mathbf{tv}_{5053}^{\mathrm{PT}} \end{pmatrix}^{\mathrm{T}}$$

and a set of topic vectors $\{\mathbf{tv}^{\mathrm{Q1}}, \mathbf{tv}^{\mathrm{Q2}}, \cdots, \mathbf{tv}^{\mathrm{Q90}}\}$ corresponding to each question in Q.

2.2.5. Relating Lecture and Question Embeddings

Embedding both videos and all the questions using a common model enabled us to compare the content from different moments of one lecture, compare the content across videos, and estimate potential associations between specific questions and specific moments of video. This allowed us to pursue several approaches to confirm the validity of our embeddings with respect to their conceptual content.

First, we reduced the dimensionality of our embeddings to a visualizable format. We did so using PCA to project the embeddings onto their first three principal components and plotting the resulting 3-dimensional embeddings (see Fig. 3.1).

Furthermore, we calculated the within-topic mean and within-topic variance of each topic across each lecture trajectory and across all questions that belonged to the same question set (see Fig. 3.2a and 3.2c). In particular, given a $t \times k$ lecture

trajectory $\mathbf{T}_{\mathrm{L}}$, we computed the column-wise mean vector

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{T}_{\mathrm{L}}} &= \begin{pmatrix} \bar{w}_1^{\mathrm{L}} & \bar{w}_2^{\mathrm{L}} & \cdots & \bar{w}_k^{\mathrm{L}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{t}\sum_{i=1}^{t} w_{i,1}^{\mathrm{L}} & \frac{1}{t}\sum_{i=1}^{t} w_{i,2}^{\mathrm{L}} & \cdots & \frac{1}{t}\sum_{i=1}^{t} w_{i,k}^{\mathrm{L}} \end{pmatrix}\end{aligned}$$

and the column-wise standard deviation vector

$$\begin{aligned}\boldsymbol{\sigma}_{\mathbf{T}_{\mathrm{L}}} &= \begin{pmatrix} \sigma_{\mathbf{w}_1}^{\mathrm{L}} & \sigma_{\mathbf{w}_2}^{\mathrm{L}} & \cdots & \sigma_{\mathbf{w}_k}^{\mathrm{L}} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\frac{(\sum_{i=1}^{t} w_{i,1}^{\mathrm{L}} - \bar{w}_1^{\mathrm{L}})^2}{t}} & \sqrt{\frac{(\sum_{i=1}^{t} w_{i,2}^{\mathrm{L}} - \bar{w}_2^{\mathrm{L}})^2}{t}} & \cdots & \sqrt{\frac{(\sum_{i=1}^{t} w_{i,k}^{\mathrm{L}} - \bar{w}_k^{\mathrm{L}})^2}{t}} \end{pmatrix}.\end{aligned}$$

Similarly, given a question set Q_L, we computed the across-question mean vector

$$\begin{aligned}\boldsymbol{\mu}_{Q_{\mathrm{L}}} &= \begin{pmatrix} \bar{w}_1^{Q_{\mathrm{L}}} & \bar{w}_2^{Q_{\mathrm{L}}} & \cdots & \bar{w}_k^{Q_{\mathrm{L}}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{|Q_{\mathrm{L}}|}\sum_{\mathrm{Q}i \in Q_{\mathrm{L}}} w_1^{\mathrm{Q}i} & \frac{1}{|Q_{\mathrm{L}}|}\sum_{\mathrm{Q}i \in Q_{\mathrm{L}}} w_2^{\mathrm{Q}i} & \cdots & \frac{1}{|Q_{\mathrm{L}}|}\sum_{\mathrm{Q}i \in Q_{\mathrm{L}}} w_k^{\mathrm{Q}i} \end{pmatrix}\end{aligned}$$

and the across-question standard deviation vector

$$\begin{aligned}\boldsymbol{\sigma}_{Q_{\mathrm{L}}} &= \begin{pmatrix} \sigma_{\mathbf{w}_1}^{Q_{\mathrm{L}}} & \sigma_{\mathbf{w}_2}^{Q_{\mathrm{L}}} & \cdots & \sigma_{\mathbf{w}_k}^{Q_{\mathrm{L}}} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\frac{(\sum_{\mathrm{Q}i \in Q_{\mathrm{L}}} w_1^{\mathrm{Q}i} - \bar{w}_1)^2}{|Q_{\mathrm{L}}|}} & \sqrt{\frac{(\sum_{\mathrm{Q}i \in Q_{\mathrm{L}}} w_2^{\mathrm{Q}i} - \bar{w}_2)^2}{|Q_{\mathrm{L}}|}} & \cdots \sqrt{\frac{(\sum_{\mathrm{Q}i \in Q_{\mathrm{L}}} w_k^{\mathrm{Q}i} - \bar{w}_k)^2}{|Q_{\mathrm{L}}|}} \end{pmatrix}.\end{aligned}$$

In addition, we computed the Pearson correlation coefficient between the within-topic mean and variance of each lecture and the within-topic mean and variance of each question set (see Fig. 3.2b and 3.2d).

Finally, we calculated the similarity between each moment in the lectures and each question in their corresponding question set by computing the distance correlation between each moment in the lecture trajectory and the topic vector of each question

using the SciPy method `cdist`. This yielded a similarity vector

$$\mathbf{sim}_{\mathrm{Q}i,\mathrm{L}} = \begin{pmatrix} s_1^{\mathrm{Q}i,\mathrm{L}} & s_2^{\mathrm{Q}i,\mathrm{L}} & \cdots & s_t^{\mathrm{Q}i,\mathrm{L}} \end{pmatrix}^{\mathrm{T}} \tag{2.1}$$

between each question's topic vector $\mathbf{tv}^{\mathrm{Q}i}$, $\mathrm{Q}i \in Q_{\mathrm{L}}$, and the corresponding lecture trajectory $\mathbf{T}_{\mathrm{L}}$, for $\mathrm{L} \in \{\mathrm{EF}, \mathrm{PT}\}$. In particular, for $j \in \{1, 2, \cdots, t\}$

$$s_j^{\mathrm{Q}i,\mathrm{L}} = \frac{(\mathbf{tv}_j^{\mathrm{L}} - \overline{\mathrm{tv}}_j^{\mathrm{L}}) \cdot (\mathbf{tv}^{\mathrm{Q}i} - \overline{\mathrm{tv}}^{\mathrm{Q}i})}{\|\|(\mathbf{tv}_j^{\mathrm{L}} - \overline{\mathrm{tv}}_j^{\mathrm{L}})\|\|^2 \|\|(\mathbf{tv}^{\mathrm{Q}i} - \overline{\mathrm{tv}}^{\mathrm{Q}i})\|\|^2},$$

where

$$\overline{\mathrm{tv}}_j^{\mathrm{L}} = \frac{1}{k} \sum_{l=1}^{k} w_{j,l}^{\mathrm{L}}$$

and

$$\overline{\mathrm{tv}}^{\mathrm{Q}i} = \frac{1}{k} \sum_{l=1}^{k} w_{j,l}^{\mathrm{Q}i}.$$

We can use the resulting similarity scores as an estimate of the relevance of the content presented at a particular moment in the lecture to a specific question (see Fig. 3.3).

Section 2.3

EEG Preprocessing

2.3.1. Data Cleaning

We performed preprocessing of the 42 valid data sets to identify, remove, and interpolate bad channels, DC signal noise, and eye movement artifacts. In general, we followed the approach proposed by Delorme [2023] in keeping preprocessing minimal. For each data set, we performed the following preprocessing steps.

First, we retrieved the two raw EEG traces corresponding to the window of time during which the participant was watching each lecture video and bandpass filtered each trace at 0.5 Hz and 60 Hz to remove low-frequency drifts and high-frequency

muscle-related activity, respectively. To do so, we used a finite input response (FIR) filter design using the window method implemented in the `filter` method of the MNE Python package (Larson et al. [2024]) with default arguments (see Appendix A, 5.1).

We then identified bad channels. We considered a channel bad if it contained missing values or if its standard deviation or median absolute deviation was below a threshold of 1e-15 V. We also identified bad channels using a random sample consensus approach (RANSAC) as proposed by Fischler and Bolles [1987] (see Appendix A, 5.2.2). To do so, we used the PyPREP Python package Appelhoff et al. [2023] methods `find_bad_by_nan_flat` with default arguments and `find_bad_by_ransac` with a correlation threshold of 0.85. We chose this threshold to mirror the EEGLAB Delorme and Makeig [2004] RANSAC implementation `clean_channels` as proposed by Delorme [2023]. If we identified bad channels, we interpolated them using spherical spline interpolation (Perrin et al. [1989]) implemented in the MNE method `interpolate_bads` with default arguments.

Finally, we performed independent component analysis (ICA) with the Picard method to identify eye movement artifacts (see Appendix A, 5.2.3). We used the MNE implementation `ica` configured to find the smallest number of independent components to explain at least 99% of the variance. We then identified and removed components related to horizontal eye movement based on the signal of the HEOG channel and components related to vertical eye movement and blinks using a hand-selected template as proposed by Bialas [2024] (see Fig. 3.4).

2.3.2. Generation of Neural Features

After all data sets had been cleaned, we prepared them for further analysis. For each data set, we split the trace of each channel into five frequency bands corresponding to the five major types of brain waves (δ: 2-4 Hz, θ: 4-8 Hz, α: 8-12 Hz, β: 12-

30 Hz, γ: 30-60 Hz) and calculated the corresponding narrowband power using the Hilbert transform implemented in the SciPy method `hilbert`. This produced five neural features for each channel or 310 features per participant, corresponding to 62 channels, in total. This number differs from the initially reported 64 channels, since the signal of one electrode (located at the right mastoid) is used as a reference channel at the time of recording and the signal of a second electrode (located at the left mastoid) is provided as an additional rereference channel. Notably, we omitted rereferencing due to its potential negative impact on downstream analyzes found by Delorme [2023].

Finally, we downsampled the neural features from 500 Hz to 10 Hz using the MNE method `resample` to alleviate the computational burden of subsequent analyzes and resampled the resulting features to exactly match the length of their corresponding lecture video using the Scipy method `resample`. The latter step was necessary to normalize the feature length to the length of the embedded lecture trajectory. This process yielded one $t \times 310$ neural feature matrix $\mathbf{E}^{\mathrm{L}}_{\mathrm{P}i,\mathbf{H}}$ for a given participant $\mathrm{P}i \in P$ and a given lecture L of length t, where $H = \{\delta, \theta, \alpha, \beta, \gamma\}$.

Section 2.4

Inter-Subject Functional Correlations

2.4.1. Definition of ISFC

Suppose that we are given a set of N neural feature matrices $F = \{\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_N\}$, where some recording $\mathbf{F}_k \in F$ has m neural features each containing t samples. Let

$$\mathbf{F}_k = \begin{pmatrix} \mathbf{c}^k_1 & \mathbf{c}^k_2 & \cdots & \mathbf{c}^k_m \end{pmatrix}^{\mathrm{T}}.$$

Further, let

$$\mathbf{F}_{\bar{\setminus k}} = \left(\mathbf{c_1}^{\bar{\setminus k}} \quad \mathbf{c_2}^{\bar{\setminus k}} \quad \cdots \quad \mathbf{c_m}^{\bar{\setminus k}} \right)^{\mathrm{T}} = \frac{1}{N-1} \sum_{i \neq k} \mathbf{F}_i$$

be the average signal of the set $F \backslash \{\mathbf{F}_k\}$. Then, the ISFC matrix of $\mathbf{F}_k$ is defined as

$$\mathbf{ISFC}_k = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,m} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \cdots & r_{m,m} \end{pmatrix}, \tag{2.2}$$

where $r_{i,j}$ is the Pearson correlation coefficient between $\mathbf{c}_i^k$ and $\mathbf{c}_i^{\bar{\setminus k}}$.

2.4.2. ISFC Generation

Instead of a global average ISFC matrix, our objective was to generate one ISFC matrix for each timepoint to serve as an estimate of a participant's understanding of the conceptual content presented at that moment in a lecture. Hence, we followed the sliding-window approach proposed by Simony et al. [2016], which modifies Eq. 2.2 by considering only a finite window of size w when computing the ISFC matrix at any given timepoint. In particular, given our preprocessed set of neural feature matrices

$$E = \{\mathbf{E}^{\mathrm{EF}}_{\mathrm{P}i,\mathbf{H}}, \mathbf{E}^{\mathrm{PT}}_{\mathrm{P}i,\mathbf{H}} : 1 \leq i \leq 42\}$$

we defined the ISFC matrix of recording $\mathbf{E}^{\mathrm{L}}_{\mathrm{P}i,\mathbf{H}}$ at timepoint t as

$$\mathbf{ISFC}^{\mathrm{L}}_{\mathrm{P}i,\mathbf{H};t} =$$

$$\left(\begin{array}{cccc|c|cccc}
r^{\mathrm{P}i,\mathrm{L}}_{1,1;t} & r^{\mathrm{P}i,\mathrm{L}}_{1,2;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{1,m;t;\boldsymbol{\eta}_m} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{1,k(m-1);t} & r^{\mathrm{P}i,\mathrm{L}}_{1,k(m-1)+1;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{1,km;t;\boldsymbol{\eta}_m} \\
r^{\mathrm{P}i,\mathrm{L}}_{2,1;t} & r^{\mathrm{P}i,\mathrm{L}}_{2,2;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{2,m;t;\boldsymbol{\eta}_m} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{2,k(m-1);t} & r^{\mathrm{P}i,\mathrm{L}}_{2,k(m-1)+1;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{2,km;t;\boldsymbol{\eta}_m} \\
\vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\
r^{\mathrm{P}i,\mathrm{L}}_{m,1;t} & r^{\mathrm{P}i,\mathrm{L}}_{m,2;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{m,m;t;\boldsymbol{\eta}_m} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{m,k(m-1)} & r^{\mathrm{P}i,\mathrm{L}}_{m,k(m-1)+1;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{m,km;t;\boldsymbol{\eta}_m} \\
\hline
\vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\
\hline
r^{\mathrm{P}i,\mathrm{L}}_{k(m-1),1;t} & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1),2;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1),m;t;\boldsymbol{\eta}_m} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1),k(m-1);t} & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1),k(m-1)+1;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1),km;t;\boldsymbol{\eta}_m} \\
r^{\mathrm{P}i,\mathrm{L}}_{k(m-1)+1,1;t} & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1)+1,2;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1)+1,m;t;\boldsymbol{\eta}_m} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1)+1,k(m-1);t} & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1)+1,k(m-1)+1;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{k(m-1)+1,km;t;\boldsymbol{\eta}_m} \\
\vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\
r^{\mathrm{P}i,\mathrm{L}}_{km,1;t} & r^{\mathrm{P}i,\mathrm{L}}_{km,2;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{km,m;t;\boldsymbol{\eta}_m} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{km,k(m-1)} & r^{\mathrm{P}i,\mathrm{L}}_{km,k(m-1)+1;t;\boldsymbol{\eta}_1} & \cdots & r^{\mathrm{P}i,\mathrm{L}}_{km,km;t;\boldsymbol{\eta}_m}
\end{array}\right)$$

where $r^{\mathrm{P}i,\mathrm{L}}_{i,j;t}$ is the Pearson correlation coefficient between $\mathbf{c}^{\mathrm{P}i,\mathrm{L}}_{i,t:t+w}$ and $\mathbf{c}^{\setminus\bar{\mathrm{P}}i,\mathrm{L}}_{i,t:t+w}$, and $r^{\mathrm{P}i,\mathrm{L}}_{i,j;t;\boldsymbol{\eta}_j}$ is the Pearson correlation coefficient between $\mathbf{c}^{\mathrm{P}i,\mathrm{L}}_{i,t:t+w;\boldsymbol{\eta}_j}$ and $\mathbf{c}^{\setminus\bar{\mathrm{P}}i,\mathrm{L}}_{i,t:t+w;\boldsymbol{\eta}_j}$, for $\eta_j \in H$.

Specifically, we used a sliding window of size $w = 100$ and stride 1, corresponding to a 10-second EEG window used for the ISFC computation at each timepoint. We used the left endpoint of the window as the ISFC's timepoint to account for the amount of time the brain requires to process the conceptual content that is presented at timepoint t. Furthermore, we standardized each neural feature prior to computing ISFCs to account for variations in amplitude between recordings that could otherwise skew the average signal $\mathbf{E}^{L}_{\setminus\bar{\mathrm{P}}i,\mathbf{H}}$. This procedure yielded one ISFC matrix timeseries for each lecture per participant.

Since our final objective was to generate a scalar-valued knowledge estimate for each timepoint, we then computed six averaged ISFC values of each ISFC matrix. In particular, we computed one average ISFC value of the full ISFC matrix and one average ISFC value for each of the blocks of the ISFC matrix corresponding to the five brain wave frequency bands. To account for potential skewness of high absolute

correlation values, we performed a Fisher z-transformation (i.e., the inverse hyperbolic tangent function) before averaging. We then remapped the averages using the inverse transformation.

Thus, we obtained six scalar-valued knowledge estimate candidates that give an approximation of the knowledge each participant has of the concepts presented at each moment in Lecture 1 and Lecture 2. In particular, for a given participant $\mathrm{P}i \in P$, we generated a set of 12 knowledge estimate vectors

$$I_{\mathrm{P}i} = \{\overline{\mathbf{isfc}}^{\mathrm{L}}_{\mathrm{P}i}, \overline{\mathbf{isfc}}^{\mathrm{L}}_{\mathrm{P}i,\delta}, \overline{\mathbf{isfc}}^{\mathrm{L}}_{\mathrm{P}i,\theta}, \overline{\mathbf{isfc}}^{\mathrm{L}}_{\mathrm{P}i,\alpha}, \overline{\mathbf{isfc}}^{\mathrm{L}}_{\mathrm{P}i,\beta}, \overline{\mathbf{isfc}}^{\mathrm{L}}_{\mathrm{P}i,\gamma} : \mathrm{L} \in \{\mathrm{EF}, \mathrm{PT}\}\}.$$

Then, for all $\overline{\mathbf{isfc}}^{\mathrm{L}}_{\mathrm{P}i,\cdot} \in I_{\mathrm{P}i}$, we define

$$\overline{\mathbf{isfc}}^{\mathrm{L}}_{\mathrm{P}i,\cdot} = \begin{pmatrix} \mathrm{isfc}^{\mathrm{P}i,\mathrm{L}}_{1,\cdot} & \mathrm{isfc}^{\mathrm{P}i,\mathrm{L}}_{2,\cdot} & \cdots & \mathrm{isfc}^{\mathrm{P}i,\mathrm{L}}_{t,\cdot} \end{pmatrix}^{\mathrm{T}} \in \mathcal{M}_{t\times 1}(\mathbb{R}),$$

where t is the length of lecture L, and $\cdot$ is either empty or in H.

2.4.3. ISFC Validation Procedure

We aimed to validate our six ISFC-derived knowledge estimate candidates by making use of the participants' performance on the quizzes. Specifically, given a participant $\mathrm{P}i \in P$, we used our moment-by-moment knowledge estimate candidates in $I_{\mathrm{P}i}$ to generate six knowledge estimates corresponding to each question they had answered. In particular, given a question's topic vector $\mathbf{tv}^{\mathrm{Q}i}$, where $\mathrm{Q}i \in Q_{\mathrm{L}}$, we retrieved its similarity vector $\mathbf{sim}_{\mathrm{Q}i,\mathrm{L}}$ which we derived in Eq. 2.1. We then computed an average knowledge estimate for $\mathrm{Q}i$ weighted by $\mathbf{sim}_{\mathrm{Q}i,\mathrm{L}}$:

$$\overline{\mathrm{isfc}}_{\mathrm{P}i,\mathrm{Q}i} = \frac{\sum_{k=1}^{t} s_k^{\mathrm{Q}i,\mathrm{L}} \cdot \mathrm{isfc}^{\mathrm{P}i,\mathrm{L}}_{k,\cdot}}{\sum_{k=1}^{t} s_k^{\mathrm{Q}i,\mathrm{L}}}.$$

We computed these question knowledge estimates for each question that the participant answered after watching its corresponding lecture video. Thus, we obtained 20 question knowledge estimates for Lecture 1 and 10 question knowledge estimates for Lecture 2 for each of our six knowledge estimate candidates. We then performed several statistical analyzes to validate these estimates. First, for each participant, we performed Welch's t-test to test if the mean of the question knowledge estimates corresponding to correctly answered questions was significantly different from the mean of the question knowledge estimates corresponding to incorrectly answered questions. This procedure yielded 42 t-statistic p-value pairs for each of our six knowledge estimate candidates.

To control for potential randomness in our results, we also performed this procedure for the set questions that participants answered before watching the corresponding lecture. Since our knowledge estimation method relies on neurophysiological recordings made during lecture viewing, we hypothesize that there should be no difference in means between the question knowledge estimates corresponding to correctly answered questions and the question knowledge estimates corresponding to incorrectly answered questions.

Chapter 3

Results

Section 3.1

Evaluating Lecture and Question Embeddings

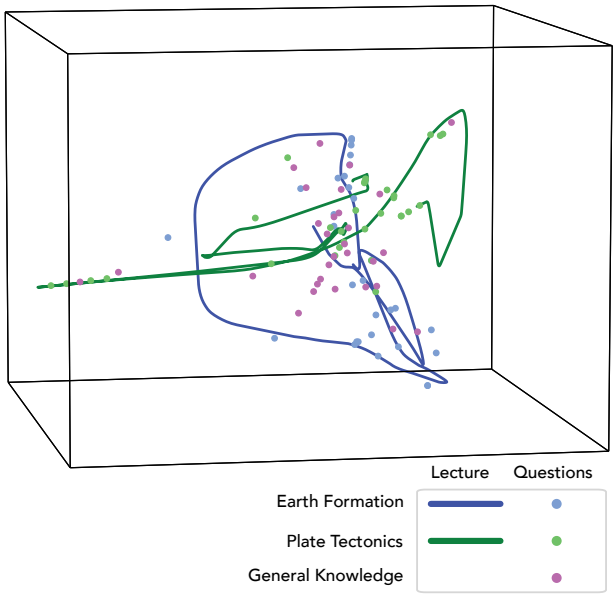
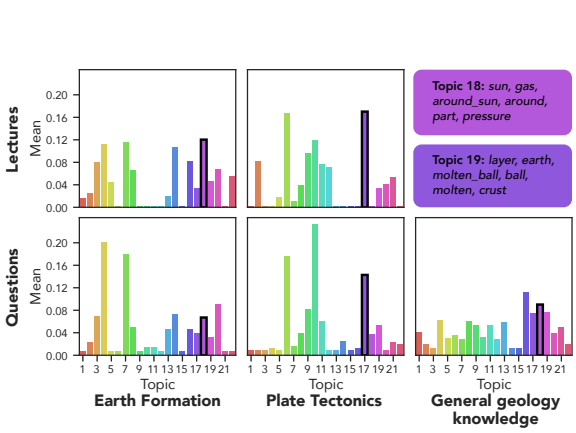


Figure 3.1: **Lecture trajectories and questions in joint topic space.** We used PCA to project the 22-dimensional embeddings of both the lecture trajectories and the questions onto their first three principal components for visualization.

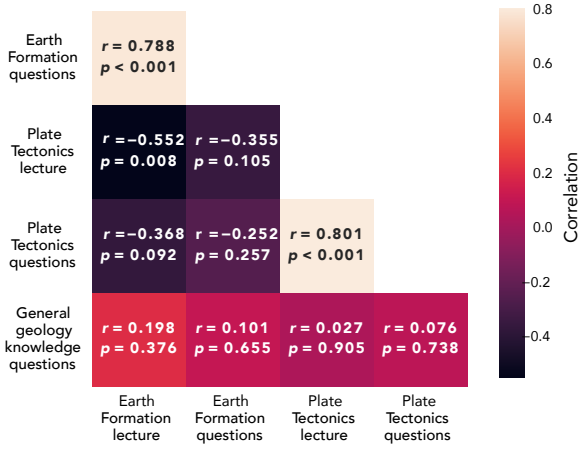
We used the results of our methodology in Section 2.2.5 to validate that the questions

from each question set aligned with their corresponding lecture and were distinct from the other lecture. First, a visual inspection of the lecture trajectories and question embeddings in Fig. 3.1 revealed an acceptable degree of separation between each lecture video and a good match between the lectures and their corresponding question set.

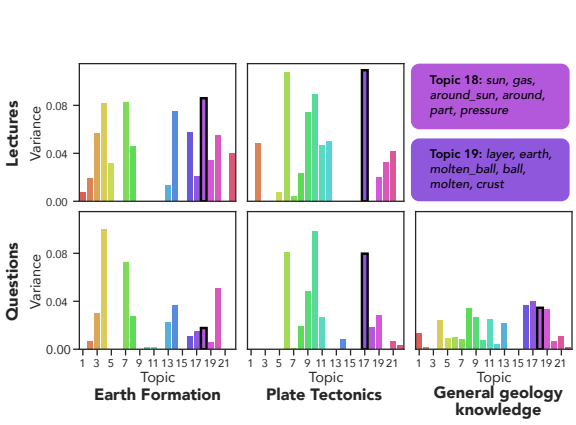
Furthermore, the within-topic means and variances between lecture-question set



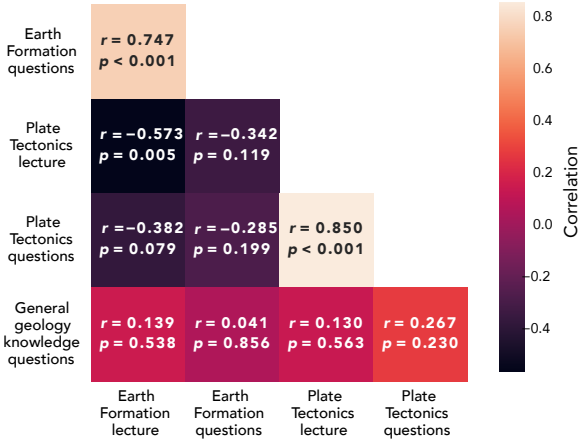
(a) Within-topic mean of each topic for lectures and question sets.



(b) Correlation between within-topic means of lecture and question set within and across lectures.



(c) Within-topic variance of each topic for lectures and question sets.



(d) Correlation between lecture and question set within and across lectures.

Figure 3.2: **Topic Mean and Variance Analyzes.** Both analyzes show a good match between lectures and their corresponding question set and good distinction between lectures.

pairs generally agreed for matching pairs and disagreed for mismatched pairs (see Figs. 3.2a and 3.2c). Additionally, the means and variances of the general knowledge question set appeared relatively uniformly distributed. We saw a similar pattern for the correlation analysis: Matching lecture-question set pairs were highly correlated, whereas mismatched pairs displayed either no correlation or negative correlation (see Figs. 3.2b and 3.2d). Overall, we observed that the lecture trajectories matched well with their respective question sets.

On the other hand, the lecture-question similarity scores in Fig. 3.3 revealed a less clearly separated relationship between individual questions and their respective lecture trajectory. While questions from question set Q_{EF} generally had well differentiated similarity scores, questions from question set Q_{PT} did not show similar levels of distinction with respect to their similarities scores.

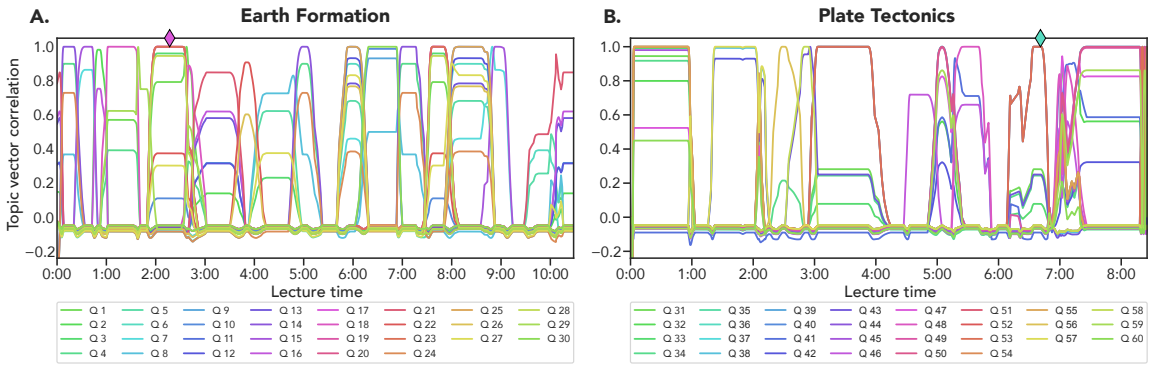


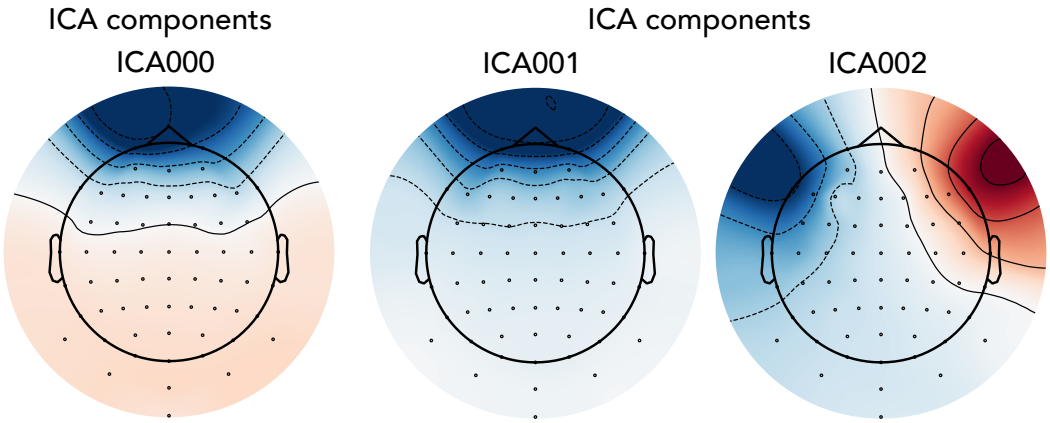
Figure 3.3: **Lecture-Question Similarity.** We used distance correlation as a measure of similarity between each moment in the lecture trajectories and each question in their corresponding question set. Question similarity scores in panel A are generally better separated than question similarity scores in panel B.

Section 3.2

ICA-Based Eye Artifact Removal

One of the greatest risks in performing ICA on EEG data is loss of information due to overly aggressive artifact removal. To verify that the independent components

selected by our ICA procedure described in Section 2.3.1 are generally associated with eye artifacts, we plotted the *scalp topographies* of the selected components and performed a manual inspection (see Fig. 3.4 for a sample). Overall, we were able to confirm that our procedure worked correctly.



(a) Hand-selected ICA template for vertical eye movements and blinks.

(b) Sample of identified ICA components related to vertical eye movements and blinks (left) and horizontal eye movements (right).

Figure 3.4: **ICA components related to eye movements.** These scalp topographies show the estimated strength of electrical activity across the scalp corresponding to components related to eye artifacts.

Section 3.3

ISFC Validation Results

The ISFC-validation procedure described in Section 2.4.3 yielded 42 t-statistic p-value pairs for each of our six knowledge estimate candidates. This gave us a distribution of t-statistics for each frequency band. Fig. 3.5 shows that no frequency band displays unambiguous results that would allow us to reject the null hypothesis.

Nevertheless, we can see that there exists a noticeable difference between the distribution of t-statistics in panels A and B, which correspond to our knowledge estimate, and the distribution of t-statistics in panels C and D, which correspond to our

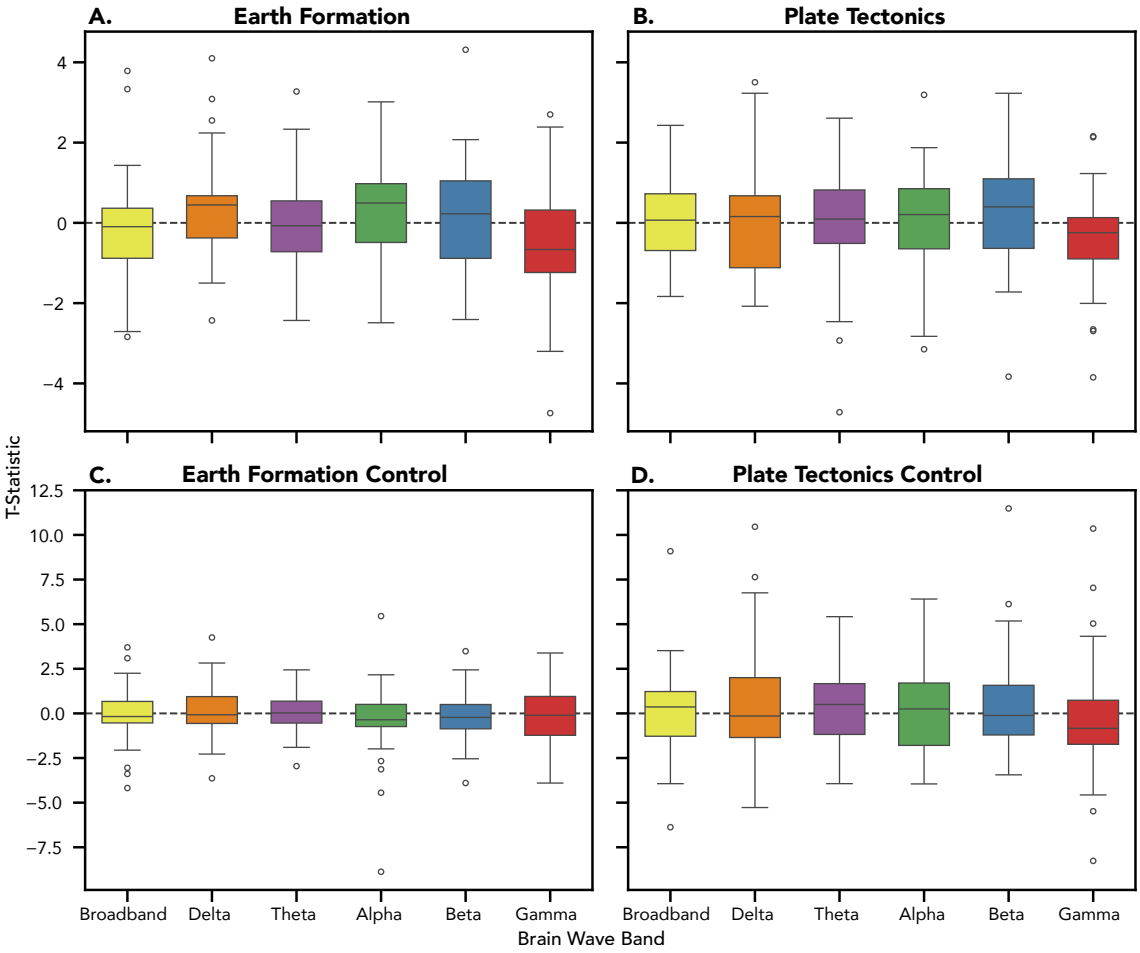
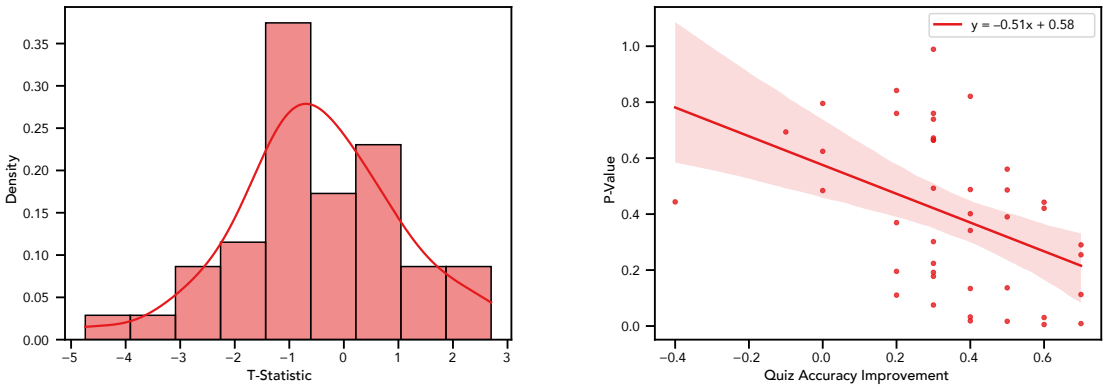


Figure 3.5: **Distribution of t-statistics by brain wave frequency band.** Each t-statistic measures the difference of means between ISFC values of correctly and incorrectly answered quiz questions. Panels A and B show the distribution of questions answered after watching the corresponding lecture. Panels C and D show the distribution of questions answered before watching the corresponding lecture.

control procedure. Specifically, the control distributions appear to be centered more closely around 0, than the distributions corresponding to our knowledge estimates.

Furthermore, we can see that the mean of the distribution of t-statistics corresponding to gamma band knowledge estimate is shifted towards the left. This is particularly apparent for Lecture 1 (see Fig. 3.6a). Lastly, we examined the relationship between the confidence of our knowledge estimates for Lecture 1 for a given participant P$i \in P$ (measured by the p-value) and their learning progress during

Lecture 1 (measured by their improvement in accuracy between Quiz 1 and Quiz 2) by fitting an OLS model. This revealed a strong linear relationship between greater learning progress and higher confidence in our knowledge estimates (see Fig. 3.6b).



(a) Distribution of t-statistics in the gamma band.

(b) OLS regression between quiz accuracy improvement and p-values

Figure 3.6: **Gamma band t-statistic and p-values for Lecture 1.** Gamma band ISFC-derived knowledge estimates may contain an informative signal (left). This signal appears to become more confident as we consider participants who learned more as judged by their quiz accuracy improvement (right).

Chapter 4

Discussion

We developed a computational approach that could provide moment-by-moment knowledge estimates of the content presented in a lecture based on neurophysiological data recorded while a participant was watching the lecture. In particular, we aimed to solve two problems: first, to quantify the conceptual content that is presented at any given point in a lecture video, and second, to use raw EEG recordings to compute a knowledge estimate of this conceptual content. To solve the first problem, we followed a topic modeling-based approach first proposed by Fitzpatrick et al. [2023]. In general, we found that this approach can generate moment-by-moment estimates of the conceptual content of a lecture and match them with the conceptual content probed by a quiz question. However, we found that the model struggled with differentiating individual questions corresponding to Lecture 2 (see Fig. 3.3). This could be due to a number of different reasons. First, a set of 30 questions might have exhausted the number of distinct concepts presented in Lecture 2, leading to conceptual overlaps between individual questions. Second, we chose a number of different hyperparameters, such as the window size of the sliding window during text preprocessing or the number of topics, which could have been ill-suited for Lecture 2.

To solve the second problem, we treated ISFC values, generated from EEG data

recorded while participants were watching the lecture videos, as estimates of knowledge and learning. Overall, we did not find conclusive results that strongly indicate the validity or invalidity of this approach. Although none of the ISFC-derived knowledge estimates allowed us to reject the null hypothesis, several of our results indicated that our knowledge estimates may contain an informative signal. We found that the distribution of t-statistics corresponding to our knowledge estimates differed from the control distribution (see Fig. 3.5). Especially gamma band knowledge estimates showed a possible difference in means between knowledge estimates corresponding to correctly and incorrectly answered questions (see Fig. 3.6a). Moreover, we found that our approach performs better (as a measure of observed p-values) for individuals who displayed greater learning progress during lecture viewing (see Fig. 3.6b).

We surmise that several factors may affect the strength of our methodology. First, the nature of our validation procedure, which is based on the ISFC responses of a given participant P$i \in P$, prevents us from using the collected quiz answers and corresponding knowledge estimates across all participants in a single statistical test. That is because we cannot assume that all participants would have the same ISFC response to a given lecture. Hence, we are constrained to each participant's individual responses (20 samples for Lecture 1 and only 10 samples for Lecture 2!) which may leave our analysis statistically underpowered. Additionally, the extremely small sample size of 10 for Lecture 2 may help explain the weakness of our findings when compared to Lecture 1. Furthermore, the presence of participants who did not learn anything while watching the lectures, for instance, since they already had strong pre-existing knowledge of the presented concepts, may have introduced a sizeable amount of noise in the data set. Since our ISFC-based approach relies on the informativity of an averaged signal (see Eq. 2.2, such noise may have negatively impacted our analysis. This is supported by our findings shown in Fig. 3.6b. We plan to investigate this

possibility in future work. We also hope to gain a better understanding of the exact electrode locations that are the most informative and pursue other computational approaches.

We hope that our analysis has taken us a step closer to establishing a method to track moment-by-moment learning that can inform models of real-world educational scenarios and how people behave in them.

Chapter 5

Appendix A

Section 5.1

Topic Modeling

5.1.1. Latent Dirichlet Allocation

Latent Dirichlet allocation (LDA) is a Bayesian hierarchical model proposed by Blei et al. [2003]. LDA assumes that documents, that is, sequences of words, can be represented as random mixtures over a fixed number of latent topics, where each topic is characterized by a distribution over words. More precisely, LDA assumes that, given a vocabulary of v words and a number of topics k, a corpus of m documents is generated according to the following process:

(a) For each document $i \in \{1, 2, \cdots, m\}$, choose $\boldsymbol{\theta}_i \sim \mathrm{Dir}_k(\alpha)$.

(b) For each topic $j \in \{1, 2, \cdots, k\}$, choose $\boldsymbol{\varphi}_j \sim \mathrm{Dir}_v(\beta)$.

(c) We note that $\boldsymbol{\theta}_i$ and $\boldsymbol{\varphi}_j$ are Dirichlet random vectors. Thus, they take points in the $(k-1)$-simplex and in the $(v-1)$-simplex, respectively. This implies that we can treat $\boldsymbol{\theta}_i$ and $\boldsymbol{\varphi}_j$ as random probability vectors.

(d) Assuming that document i contains n_i words, each word w_n in document i is generated as follows:

- Choose a topic $z_n \sim \text{Multinomial}(\boldsymbol{\theta}_i)$.
- Choose a word $w_n \sim \text{Multinomial}(\boldsymbol{\varphi}_{z_n})$.

Note that this formulation treats α and β as priors that are determined at the start of the generation process of a corpus. Thus, when fitting an LDA model to a corpus M, we are attempting to find the parameters α and β that maximize the log likelihood of the observed data M. Once we have approximated these parameters, we can represent a topic as a mixture of words and a document as a mixture of topics. Notably, a fitted LDA model can also represent unseen documents as a mixture of the topics it found during fitting.

Section 5.2

Signal Processing

5.2.1. FIR Filters

In general terms, a finite impulse response (FIR) filter is a type of discrete-time filter "based on directly approximating the desired frequency response or impulse response of the discrete-time system." Oppenheim and Schafer [2009]. In our case, we applied an FIR filter using the window method, which is the most straightforward design on an FIR filter. In particular, given an input signal $x[t]$ at time t and a finite set of filter coefficients $\{b_0, b_1, \cdots, b_{M-1}\}$, we perform the discrete-time convolution

$$y[t] = \sum_{k=0}^{M-1} b_k x[t-k] \tag{5.1}$$

to calculate our filtered signal $y[t]$. We can then determine the frequency response of our filter by examining its impulse response, that is, we let

$$x[t] = \delta[t] = \begin{cases} 0 & \text{if } t \neq 0 \\ & \\ 1 & \text{if } t = 0 \end{cases}.$$

We also call $\delta[t]$ the impulse signal and we can think of it as the discrete equivalent of the unit impulse function, that is, the Dirac delta function $\delta(t)$. In particular, we make use of two properties of the unit impulse function. First, the unit impulse function has an infinite frequency bandwidth Weisstein. Second, given an arbitrary real-valued function f,

$$f(t) = \int_{-\infty}^{\infty} f(s)\delta(t-s)ds \text{ Grami [2016].}$$

Thus, Eq. 5.1 with $x[t] = \delta[t]$ yields

$$y[t] = b_t.$$

We commonly call $h[t] = y[t] = b_t$ the impulse response of our FIR filter. Finally, we can determine the discrete frequency response of our FIR filter by applying the discrete-time Fourier transform

$$H[e^{j\omega}] = \sum_{k=0}^{M-1} h[t]e^{-j\omega k}.$$

5.2.2. RANSAC

Random sample consensus (RANSAC) is a general approach to fit a statistical model to experimental data that contains a significant percentage of outliers first proposed

by Fischler and Bolles [1987]. Bigdely-Shamlo et al. [2015] adopted RANSAC as a method of detecting corrupt EEG channels. Specifically, RANSAC works by randomly selecting a subset of channels as "inliers." Data from all channels is then interpolated using only these inliers. This process is repeated n times to generate n separate time series for each channel. The correlation between the median of these n series (calculated at each time point) and the original data is then computed. If the correlation falls below a set threshold, the sensor is flagged as an outlier and is considered corrupted.

5.2.3. Independent Component Analysis

Independent component analysis (ICA) is a technique that aims to separate an observed mixed signal into maximally independent additive subcomponents (Langlois et al. [2010]). In particular, assume that we have an observed random vector of n linear mixtures

$$\mathbf{x} = \Big(x_1, \, x_2, \, \cdots \, , \, x_n \Big)^{\mathrm{T}}$$

where for all $i \in \{1, \, 2, \, \cdots \, , \, n\}$, x_i is the mixture of n independent elements of a latent random vector $\mathbf{s} = \Big(s_1, \, s_2, \, \cdots \, , \, s_n \Big)^{\mathrm{T}}$, that is,

$$x_i = a_{i,1} s_1 + a_{i,2} s_2 + \cdots + a_{i,n} s_n.$$

Then,

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

where $\mathbf{A}$ is an $n \times n$ mixing matrix. The goal of ICA is to approximate $\mathbf{s}$ by approximating the inverse of the mixing matrix $\mathbf{W} = \mathbf{A}^{-1}$. Once we have an approximation of $\mathbf{W}$, we can easily compute

$$\mathbf{s} = \mathbf{W}\mathbf{x}.$$

Chapter 6

Appendix B

Section 6.1

Code and Data Availability

The entire code base is available here.

Section 6.2

Identified Topics

Table 6.1: Top 10 words for each topic identified by the LDA model

Topic 1	earth	year	time	time_clock	clock	ago	think	crust	probably	theia
Topic 2	plate	mantle	solid	earth	think	crust	form	glance	layer	kind
Topic 3	plate	travel	supernova	year	light	shockwave	speed	gas	mantle	speed_light
Topic 4	form	year	ago	supernova	time	period	earth	seem	billion	billion_year
Topic 5	ball	molten	magma	crust	earth	molten_ball	giant	rock	still	look
Topic 6	fluid	mantle	kind	layer	still	lithosphere	magma	part	property	asthenosphere
Topic 7	period	hadean	earth	eon	mantle	kind	hadean_eon	molten	time	distinctive
Topic 8	earth	supernova	layer	look	different	surface_earth	surface	magma	happen	formation
Topic 9	plate	move	relative	move_relative	move_away	away	new	land	pacific_plate	new_land
Topic 10	plate	rigid	move	pacific	pacific_plate	bunch	bunch_rigid	relative	move_relative	movement
Topic 11	earth	core	mantle	solid	plate	magma	kind	outer	layer	look
Topic 12	solid	plastic	kind	deformable	stuff	melt	mantle	plastic_solid	magma	little
Topic 13	plate	supernova	year	picture	rigid	form	move	happen	move_relative	relative
Topic 14	earth	happen	eventual	become	planet	supernova	little	formation	splash	probably
Topic 15	plate	form	travel	new_land	new	land	away	eon	move	light
Topic 16	earth	form	think	process	happen	modern_earth	modern	hadean	another	eventual
Topic 17	crust	mantle	layer	solid	different	earth	blue	division	form	core
Topic 18	sun	gas	around_sun	around	part	pressure	kind	little	start	happen
Topic 19	layer	earth	molten_ball	ball	molten	crust	real	lithosphere	asthenosphere	whole
Topic 20	plate	splash	earth	molten	orbit	splash_orbit	theia	relative	big	america
Topic 21	fluid	much	behave_fluid	behave	viscid	deformable	move	associate	thick	slow
Topic 22	earth	fluid	think	kind	blow	glance	think_glance	molten	layer	deformable

Bibliography

Stefan Appelhoff, Austin J. Hurst, Aamna Lawrence, Adam Li, Yorguin José Mantilla Ramos, Christian O'Reilly, Liang Xiang, Jonte Dancker, Mathieu Scheltienne, and Ole Bialas. PyPREP: A Python implementation of the preprocessing pipeline (PREP) for EEG data., October 2023. URL https://doi.org/10.5281/zenodo.10047462.

Ole Bialas. Eeg preprocessing ii: eye-artifacts, repairing and rejecting, February 2024. URL https://olebialas.github.io/posts/eeg_preprocessing2/.

Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A. Robbins. The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Frontiers in Neuroinformatics*, 9, 2015. ISSN 1662-5196. doi: 10.3389/fninf.2015.00016. URL https://www.frontiersin.org/articles/10.3389/fninf.2015.00016.

Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009. ISBN 978-0-596-51649-9. URL https://www.nltk.org/book/.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022, mar 2003. ISSN 1532-4435. URL https://dl.acm.org/doi/10.5555/944919.944937.

Arnaud Delorme. EEG is better left alone. *Scientific Reports*, 13(1):2372, February 2023. URL https://doi.org/10.1038/s41598-023-27528-0.

Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004. ISSN 0165-0270. URL https://doi.org/10.1016/j.jneumeth.2003.10.009.

Martin A Fischler and Robert C Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In Martin A Fischler and Oscar Firschein, editors, *Readings in Computer Vision*, pages 726–740. Morgan Kaufmann, San Francisco (CA), January 1987. URL https://doi.org/10.1145/358669.358692.

Paxton C Fitzpatrick, Andrew C Heusser, and Jeremy R Manning. Text embedding models yield high-resolution insights into conceptual knowledge from short multiple-choice quizzes. February 2023. URL https://doi.org/10.31234/osf.io/dh3q2.

Ali Grami. Chapter 3 - signals, systems, and spectral analysis. In Ali Grami, editor, *Introduction to Digital Communications*, pages 41–150. Academic Press, Boston, January 2016. ISBN 978-0-12-407682-2. URL https://doi.org/10.1016/C2012-0-06171-6.

Uri Hasson, Yuval Nir, Ifat Levy, Galit Fuhrmann, and Rafael Malach. Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664):1634–1640, March 2004. URL https://doi.org/10.1126/science.1089506.

Dominic Langlois, Sylvain Chartier, and Dominique Gosselin. An introduction to independent component analysis: Infomax and fastica algorithms. *Tutorials in*

Quantitative Methods for Psychology, 6(1):31–38, 2010. doi: 10.20982/tqmp.06.1.p031. URL http://www.tqmp.org/RegularArticles/vol06-1/p031/p031.pdf.

Eric Larson et al. Mne-python, April 2024. URL https://doi.org/10.5281/zenodo.10999175.

Alan V. Oppenheim and Ronald W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall Press, USA, 3rd edition, 2009. ISBN 0131988425.

F Perrin, J Pernier, O Bertrand, and J F Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol*, 72(2):184–187, February 1989. URL https://doi.org/10.1016/0013-4694(89)90180-6.

Andreas Trier Poulsen, Simon Kamronn, Jacek Dmochowski, Lucas C Parra, and Lars Kai Hansen. EEG in the classroom: Synchronised neural recordings during video presentation. *Sci. Rep.*, 7(1), March 2017. URL https://doi.org/10.1038/srep43916.

Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. URL http://is.muni.cz/publication/884893/en.

Erez Simony, Christopher J Honey, Janice Chen, Olga Lositsky, Yaara Yeshurun, Ami Wiesel, and Uri Hasson. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.*, 7(1):12141, November 2016. URL https://doi.org/10.1038/ncomms12141.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman,

Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. URL `https://doi.org/10.1038/s41592-019-0686-2`.

Eric W. Weisstein. *Delta Function*. MathWorld–A Wolfram Web Resource. URL `https://mathworld.wolfram.com/DeltaFunction.html`.