Analysis methods for naturalistic data

Emily Finn, PhD Co Organizer
National Institute of Health
Bethesda, MD
United States

Jeremy Manning, PhD Co Organizer
Dartmouth College
Psychological and Brain Sciences
Hanover, NH
United States

Tor Wager, PhD Co Organizer
Dartmouth College
Hanover, NH
United States

Luke Chang, PhD Organizer
Dartmouth College
Psychological & Brain Sciences
Hanover, NH
United States

This topic is timely and important because many researchers who come from resting-state or traditional task backgrounds face challenges in adapting familiar tools (e.g., functional connectivity, general linear models) for use with naturalistic data, and/or incorporating methods specific to naturalistic data that leverage unique features of these data (e.g., inter-subject, state changes, etc.).

Because naturalistic paradigms allow researchers to probe multiplexed signals across the hierarchy of neural systems—from low-level sensory processing up to social cognition—with a single acquisition, these data have high potential for reuse. Encouragingly, many authors of naturalistic imaging studies have made their data public, and some large-scale data collection efforts have begun to include naturalistic paradigms (e.g., Human Connectome Project 7T movie-watching data, Child Mind Institute Healthy Brain Network initiative). Yet for researchers to take full advantage of the richness of information in these data will require additional methods education.

The desired learning outcomes for this educational session are: 1) participants will appreciate the landscape of potential approaches to analyzing naturalistic data; 2) participants will understand how to leverage unique features of naturalistic data to conduct analyses that would not be possible with resting-state or traditional task-based designs; 3) participants will gain hands-on experience with cutting-edge Python-based tools for naturalistic data analysis, and in several cases learn directly from the core developers of these software packages; 4) participants will engage in Q&A/brainstorming sessions with experts to help them understand how to apply these methods to their own data; 5) participants will understand the need for rigor and reproducibility in the context of the heightened degrees of freedom that come along with naturalistic data; and 6) will learn concrete strategies for enforcing these principles in their own work.

Objective

1) The audience will learn innovative methods for analyzing naturalistic data to gain insights beyond what is possible with resting-state or traditional task-based designs, including both the theory behind these methods and the details of how they are implemented.

2) The audience will gain hands-on experience with cutting-edge tools for performing naturalistic data analysis (e.g., inter-subject correlation and functional connectivity, encoding and decoding models, inter-subject representational similarity analysis, automatic annotations) via tutorials using Python-based materials, which they can then take back to their labs and adapt for use with their own data.

Target Audience

Our target audience is anyone working with naturalistic datasets. We think researchers with a diverse range of backgrounds will benefit from our course ranging from intermediate to expert levels of analysis experience. Our hands on analysis demonstrations will be in the Python programming language and will be able to run on individuals' laptops. No prior experience with Python or programming will be strictly necessary, although researchers with some programming experience will likely find it easier to follow along.

Presentations

Quantifying interindividual similarity of brain and behavior in space and time

Intersubject similarity is a general group of methods for addressing similarity of brain activity, physiology and behavior. It also reflects interindividual differences, such as level of friendship or shared understanding of the external world, and is affected by expertise as well psychiatric disorders. Through examples, this lecture will cover basics of intersubject correlations and extend those to address the dynamics of similarity of brain networks in time using methods such as phase synchronization. Finally, the lecture will cover how to address differences in the levels of intersubject similarity and how these methods can be used to draw links between the similarity structure of subjective experience, behavior, and underlying brain activity.

Presenter

Juha Lahnakoski, D.Sc. (Tech.), Forschungszentrum Jülich Düsseldorf, International
Germany

Decoding narrative content from fMRI responses

Despite growing interest in the process of narrative comprehension, the neural mechanisms of how the brain builds causal narrative structures that unfold over time are still largely unknown. In this presentation, I will discuss examples of how to examine the cognitive processes of story comprehension by combining fMRI data collected during movie-watching and crowdsourcing measures of comprehension with machine learning analysis based on a word embedding model. First, I will discuss how the cognitive states that reflect changing levels of story understanding can be predicted by the patterns of time-resolved functional connectivity. Next, using behavioral measures of causal relatedness of scenes and natural language annotations, I will discuss how fMRI responses of causally related past events can be used for improving decoding of narrative contents of the movie. Our results suggest that the brain constructs coherent narratives by integrating information from the causally related past into the present moment. The analysis methods discussed in this presentation will be of interest to researchers who aim to (i) understand the human memory process in ecologically valid, naturalistic settings, and (ii) implement natural language processing to explain high-level cognitive processes.

Presenter

Won mok Shim, PhD, Sungkyunkswan University Seoul, International
Korea, Republic of

Studying the cortical hierarchy of event segmentation

A key component of the way humans think about and remember experiences is that they are segmented into discrete events. This segmentation is important for our understanding of an experience and has implications for memory and future learning. Most of the knowledge about event segmentation is based on behavioral work, but recently developed data-driven approaches allow us to use neuroimaging data of movie stimuli to study how different brain regions segment experiences into events. These methods have revealed that humans segment experience on different temporal scales across the cortical hierarchy, with short events in early sensory regions and long periods of information integration in higher level multimodal areas. In this lecture, I will explain a simple greedy search algorithm that can be used to identify event boundaries in neuroimaging data. I will then show how we can combine this approach with functional connectivity methods to investigate how event boundaries are shared between brain regions at different levels of the cortical hierarchy. Finally, I will illustrate how we can use these tools to study individual differences in event segmentation, specifically in relation to healthy aging.

Presenter

Linda Geerligs, PhD, Donders Institute Nijmegen, Gelderland
Netherlands

Modeling fMRI responses to learning, perception, and prediction with natural text

Studies investigating how knowledge of schematic sequences can influence perception and prediction traditionally use simple associations between artificial stimuli. In my lab we are leveraging recent advancements in machine learning to generate sets of stimuli with novel schematic structure that are naturalistic and engaging, capturing the complexity of extracting structure from realistic experiences. I will discuss how we developed a system that can generate a near-infinite set of novel but human-like poems, all of which progress through a pre-specified sequence of semantic topics. I will then describe how we can model the neural dynamics occurring as subjects listen to this schematically-structured text using a recurrent neural network, which has been trained on this large corpus to predict upcoming words. We find that the model automatically builds representations of contextual information present at different timescales in the poetry, such as the grammatical structure of the current line, the rhyme scheme, and the semantic sequence of the stanzas. The model can also be used to generate hypotheses about predictive coding signals in the brain, since it can quantify the degree to which upcoming words are predictable from the schematic or grammatical context. Finally, I will discuss how to apply event segmentation models to the fMRI data to identify transitions between semantic topics and predict moments during which the hippocampus will be engaged to encode episodic memories.

Presenter

Christopher Baldassano, PhD, Columbia University New York, NY
United States

High-level cognition during story listening is reflected in high-order dynamic correlations in neural activity patterns

Our thoughts arise from coordinated patterns of interactions between brain structures that change with our ongoing experiences. High-order dynamic correlations in neural activity patterns reflect different subgraphs of the brain's connectome that display homologous lower-level dynamic correlations. We tested the hypothesis that high-level cognition is supported by high-order dynamic correlations in brain activity patterns. We developed an approach to estimating high-order dynamic correlations in timeseries data, and we applied the approach to neuroimaging data collected as human participants either listened to a ten-minute story, listened to a temporally scrambled version of the story, or underwent a resting state scan. We trained across-participant pattern classifiers to decode (in held-out data) when in the session each neural activity snapshot was collected. We found that classifiers trained to decode from high-order dynamic correlations yielded the best performance on data collected as participants listened to the (unscrambled) story. By contrast, classifiers trained to decode data from scrambled versions of the story or during the resting state scan yielded the best performance when they were trained using first-order dynamic correlations or non-correlational activity patterns. We suggest that as our thoughts become more complex, they are supported by higher-order patterns of dynamic network interactions throughout the brain.

Presenter

Jeremy Manning, PhD, Dartmouth College
Psychological and Brain Sciences
Hanover, NH
United States

Examining group differences in brain responses to naturalistic stimuli

Naturalistic stimuli are complex and can lead to divergent interpretations between different individuals. For example, two people that watch a movie about veganism, while one is vegan, and the other is a meat-lover, may interpret the same movie very differently, based on their beliefs and emotions. How is this reflected in their brain responses? If one is interested in shared brain response to the movie, they may use Inter-Subject-Correlation (ISC) between these individuals. Yet what if one is interested in testing for regions that their response varies as a function of how the vegan and the meat-lover interpret the movie? In this talk I will discuss a method that is based on Euclidean distance to measure differences in the brain response to naturalistic stimuli such as movies and stories between two groups.

Presenter

Yaara Yeshurun, PhD, Tel Aviv University Tel Aviv, International
Israel

Using neural responses to naturalistic stimuli to predict social outcomes

Although much work studying fMRI responses to naturalistic stimuli has focused on commonalities across people, recent work has shown that experimentally manipulated shifts in the goals and contextual information provided to participants can powerfully shape the relative similarity of their fMRI responses to naturalistic stimuli. In addition, a growing body of literature demonstrates that relative differences in fMRI response time series evoked during naturalistic stimulation are sensitive to intrinsic individual differences in how people spontaneously process the world around them. In this talk, I will discuss recent work in which endogenous individual differences in neural responses to a wide range of naturalistic stimuli were used to successfully predict socio-behavioral outcomes, including the positions that people occupy in their real-world social networks. For example, I will discuss studies in which the relative proximity between individuals in social ties could be predicted by the relative similarity of their neural responses to naturalistic stimuli. In addition to presenting the results of recent studies predicting social network data from naturalistic neuroimaging data, I will also discuss the general promise of naturalistic neuroimaging methods for predicting social outcomes, as well as practical issues related to study design, data analytic choices, and interpretation.

Presenter

Carolyn Parkinson, PhD, University of California Los Angeles Los Angeles, CA
United States

Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging

Many powerful analysis methods for naturalistic data leverage responses that are shared across people, but individuals also exhibit idiosyncrasies in brain activity during naturalistic stimulation that may relate to their intrinsic traits, biases, and tendencies. I will review recent work using naturalistic stimuli to study individual differences, and advance a framework for detecting structure in idiosyncratic patterns of brain activity, or "idiosynchrony". Specifically, I will outline the technique of inter-subject representational similarity analysis (IS-RSA), including its theoretical motivation and an empirical demonstration of how it recovers brain-behavior relationships during movie watching using data from the Human Connectome Project. I will also discuss considerations for choosing stimuli to maximize sensitivity to individual differences.

Presenter

Emily Finn, PhD, National Institute of Health Bethesda, MD
United States

Automated feature extraction and analysis of naturalistic data

Characterizing neural responses to naturalistic experimental designs often requires leveraging the rich dynamic nature of the stimuli. However, creating manual annotations of stimuli can be an expensive, time-consuming, and laborious process that can limit our inferences. In this tutorial, I will introduce software developed in my laboratory to accelerate insight using naturalistic data by automating feature extraction and data analyses using open source cloud-based tools. For example, pliers is a tool developed in Python that can perform automated extraction of features from multimodal stimuli by providing a standardized interface to many different cloud-based services, including many state of the art deep learning models. Extractable features include a range of properties from low-level audiovisual information (e.g., luminance, audio envelope) to mid-level semantic tags (e.g., object labels) to high-level affective states (e.g., facial emotions). Neuroscout is a platform for sharing data and performing automated GLM based analyses using any combination of features. This scalable approach enables the rapid and efficient testing of novel theoretical hypotheses by reusing existing datasets that are intrinsically high dimensional data such as movies and audio narratives.

Presenter

Tal Yarkoni, PhD, University of Texas at Austin Austin, TX
United States