Timing Is Everything: Temporal Dynamics of BOLD Responses to Naturalistic Features in Movie Watching

Xin Di [1*], George B. Hanna [1], Bharat B. Biswal [1*]

1. Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ, 07102, USA

* Corresponding author:
Xin Di, Ph.D.
604 Fenster Hall, University Height
Newark, NJ, 07102, USA
xin.di@njit.edu

Bharat B. Biswal, Ph.D.
607 Fenster Hall, University Height
Newark, NJ, 07102, USA
bbiswal@yahoo.com

Abstract

Naturalistic stimuli, such as movies and narratives, are increasingly used in cognitive neuroscience to map cognitive and affective processes onto brain activity measured with functional MRI (fMRI). Features extracted from movies span multiple levels, from computational descriptions of visual and auditory input to physiological signals and subjective ratings. However, the temporal alignment between these features and the blood-oxygen-level-dependent (BOLD) response can vary considerably, and the commonly used canonical hemodynamic response function (HRF) with temporal derivatives may not adequately capture these delays. In this study, we analyzed three movie-watching datasets and examined a range of features, including visual luminance and contrast, auditory pitch, a physiological measure of pupil size, and subjective ratings of theory of mind. We performed cross-correlation analyses between feature time series and fMRI responses across the whole brain. Our results show that the canonical HRF captures sensory features well but misaligns with slower signals such as pupil size and subjective ratings. Notably, raw pupil size and subjective ratings exhibited stronger correspondence with delayed hemodynamic responses. Lastly, we showed that video data can be used as continuous input to measure hemodynamic response functions across brain regions. These findings highlight the critical role of temporal alignment in mapping naturalistic features to brain activity. Accounting for feature-specific delays may improve the interpretability and accuracy of fMRI studies using complex, real-world stimuli.

Keywords: Blood-oxygen-level dependent, Deconvolution, Naturalistic stimuli, Subjective rating, System response function, Timing.

1. Introduction

Naturalistic stimuli — such as movies or spoken narratives — have become increasingly popular in functional MRI (fMRI) research because they offer tools to investigate brain function under ecologically valid conditions (Di et al., 2026; Hasson et al., 2004; Vanderwal et al., 2019). Unlike traditional, highly controlled experimental paradigms, naturalistic stimuli capture the richness and temporal continuity of real-world experiences, thereby engaging a wide range of perceptual, cognitive, and affective processes simultaneously. This shift has opened new avenues for studying how the brain integrates complex, dynamic information over time.

A seminal line of work demonstrated that when multiple individuals watch the same movie or listen to the same story, their brain activity exhibits significant inter-subject correlations (ISC), revealing shared neural responses to naturalistic events (Hasson et al., 2004). Such data-driven analyses have been instrumental in identifying brain regions involved in high-level perception, attention, and social cognition (Kauppi et al., 2010). However, while these approaches effectively capture synchronized patterns of activity across individuals, they do not directly reveal which specific features of the stimulus drive these neural responses. To address this, researchers have increasingly focused on extracting quantitative features from naturalistic stimuli and mapping their temporal dynamics onto brain activity measured by fMRI (Bartels et al., 2008; Bartels and Zeki, 2004; Huth et al., 2016; Lahnakoski et al., 2012a; Raz et al., 2016).

A wide variety of stimulus-derived features can be used for such mapping. These range from low-level sensory features — such as visual luminance, contrast, and auditory pitch — to higher-level variables derived from physiological signals (e.g., pupil size) or subjective behavioral ratings (e.g., perceived emotional valence or theory of mind). In recent years, feature extraction has also benefited from advances in deep neural networks, where model-based representations capture complex visual, auditory, or semantic content (McMahon et al., 2023; Samara et al., 2025; Sohn et al., 2024). These features can be analyzed individually using general linear models (GLMs) or jointly in encoding-model frameworks to predict regional brain activity.

A critical challenge in linking these features to fMRI data lies in the temporal mismatch between stimulus features and the hemodynamic response. The fMRI signal—via the blood-oxygen-level-dependent (BOLD) effect—typically lags behind underlying neural activity by about 5–6 seconds (Friston et al., 1998; Handwerker et al., 2004). To account for this, most studies either apply a fixed delay (e.g., 4 s) or convolve the feature time-series with a canonical hemodynamic response function (HRF). Some approaches further include temporal derivatives of the HRF to allow limited flexibility in response timing (i.e., ≈1 s adjustment). However, it remains unclear whether these standard treatments are sufficient — especially when multiple types of features (each potentially having different temporal relationships to neural and vascular processes) are analyzed simultaneously.

Compounding this issue, the shape and latency of the HRF vary considerably across brain regions, across individuals, and even within the same region under different physiological conditions (Aguirre et al., 1998; Handwerker et al., 2004). Such variability further complicates efforts to accurately align continuous stimulus features with corresponding neural activity. This temporal uncertainty becomes especially pronounced when integrating physiological or behavioral measures that may not immediately reflect underlying neural processes.

For example, during movie watching, visual and auditory information flows from the stimulus to the brain, where neural activity arises and is subsequently captured by the MRI scanner through a delayed BOLD response. In parallel, pupil size may be measured using an eye-tracking camera and can reflect autonomic arousal or cognitive load — yet the temporal lag between pupil changes and corresponding BOLD responses remains uncertain (Joshi et al., 2016; Pais-Roldán et al., 2020). Similarly, participants may continuously provide subjective ratings (e.g., of affective or social states) using a joystick or other input device, but the timing of these behavioral signals relative to both neural activity and the BOLD signal is also unclear.

Together, these challenges highlight the importance of understanding and modelling temporal relationships among stimulus features, physiological measures, and fMRI signals (Figure 1). Establishing accurate temporal alignment is essential for interpreting the neural basis of complex cognitive and

affective processes in naturalistic contexts. The present study addresses this issue by systematically analyzing the timing of BOLD responses to multiple types of sensory, physiological, and behavioral features extracted from naturalistic movie-viewing data.

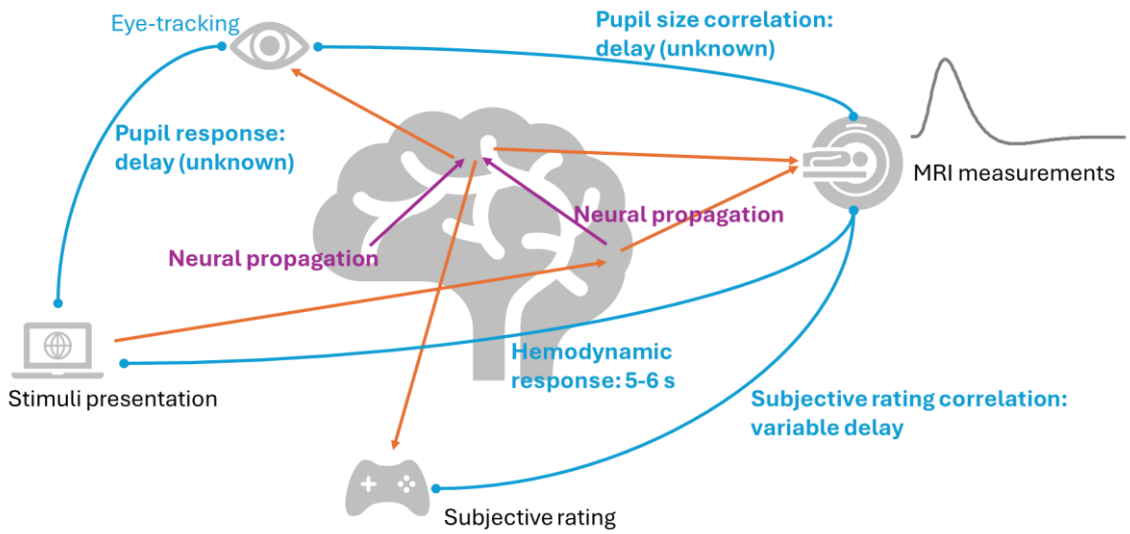

Figure 1. Conceptual diagram illustrating the temporal relationships among stimulus presentation, pupil responses, subjective ratings, and fMRI measurements during movie viewing. Blue arrows indicate measured correlations, orange arrows represent presumed causal or temporal dependencies, and purple arrows denote intrinsic and extrinsic neural propagations. The hemodynamic response is typically delayed by approximately 5–6 seconds, whereas aligning pupil and behavioral measures with BOLD signals remains challenging due to their differing temporal dynamics and unknown response delays.

Pupil size provides a valuable physiological index that bridges sensory processing, autonomic arousal and cognitive state during naturalistic viewing. Variations in pupil diameter have been linked to luminance and contrast changes in the visual input but also reflect fluctuations in cognitive load, attention, and emotional arousal (Joshi and Gold, 2020). Movie-viewing paradigms are especially well suited to capture these processes because such factors co-occur and evolve continuously, making pupil size a promising integrative marker of moment-to-moment engagement. Previous fMRI studies typically modelled pupil size by convolving its time-series with a canonical HRF before correlating it with BOLD

activity, implicitly assuming that pupil-linked neural activation follows a hemodynamic delay similar to sensory-evoked responses (Murphy et al., 2014). However, because pupil size is itself a delayed physiological signal relative to the eliciting neural activity, this conventional approach may compound delays and obscure the true temporal relationship between pupil dynamics and brain activity (Lloyd et al., 2023). Systematic examination of how the timing of pupil fluctuations aligns with BOLD responses across regions can therefore clarify whether HRF convolution is necessary — or whether raw pupil time-series may more accurately reflect brain–body coupling in naturalistic contexts.

In contrast to physiological measures, subjective ratings provide a direct behavioral window into participants' internal experiences during movie watching. Among these, theory of mind (ToM) — the capacity to infer others' emotions, thoughts, and intentions — is a core component of social cognition that naturalistic stimuli elicit particularly well (Saarimäki, 2021). Continuous ToM ratings offer a dynamic measure of how strongly a viewer engages with characters' mental states over time, enabling fine-grained mapping of social-cognitive processing in the brain. Yet, the temporal alignment of these ratings with neural activity remains uncertain. Unlike sensory features, which have relatively well-characterized delays, subjective judgments depend on higher-order cognitive integration, conscious awareness, and motor response, all of which may introduce variable and potentially substantial temporal lags. Additionally, participants may differ in how quickly they form and report ToM-related impressions. Investigating the timing of these subjective ratings with respect to BOLD activity can thus shed light on how self-reported social-cognitive experiences unfold in real time and inform future modelling strategies for linking introspective behavioral measures with neural data in naturalistic paradigms.

In the present study, we analyzed three movie-watching fMRI datasets to investigate how different types of stimulus, physiological, and behavioral features relate to brain activity over time. Specifically, we extracted multiple continuous features spanning different levels of processing: low-level sensory features (frame luminance, frame contrast, and auditory pitch), a physiological measure (pupil size), and a subjective behavioral measure (ToM ratings). We first performed cross-correlation analyses to characterize the temporal relationships between each feature and the fMRI time-series across brain

regions, thereby identifying relative delays between stimulus or behavioral dynamics and corresponding BOLD responses. In addition, we explored the use of finite impulse response (FIR) models to estimate the HRF associated with these continuous features, aiming to assess whether reliable HRF estimates can be derived in naturalistic data. Finally, we examined the spatial overlap of feature-related correlations across the three videos to evaluate the robustness and generalizability of the observed patterns.

We hypothesize that different types of features will exhibit distinct temporal relationships with the BOLD signal, reflecting differences in underlying neural and physiological processes. Specifically, low-level sensory features such as luminance and contrast are expected to correlate most strongly with early visual regions at shorter delays, whereas pupil size and ToM ratings are expected to show delayed and more diffuse correlations involving limbic and pre-frontal regions. Moreover, we expect that FIR-based HRF estimation will reveal regional variability in response latency and shape, highlighting the potential of naturalistic stimuli to probe hemodynamic dynamics across brain systems.

2. Materials and Methods

2.1. Movie-watching fMRI Data

We analyzed three movie-watching fMRI datasets previously used in our work (Di et al., 2023; Di and Biswal, 2022): two from the Healthy Brain Network (HBN) project (Alexander et al., 2017), and one from the Partly Cloudy dataset (Richardson et al., 2018).

2.1.1. Healthy Brain Network (HBN) Dataset

The HBN dataset was obtained from the project repository (http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/). For the current study, we included a subset of participants without diagnosed psychiatric disorders who underwent fMRI scanning while watching two video clips: *The Present* (3 min 21 s; Filmakademie Baden-Wuerttemberg, 2014) and a 10-minute segment from *Despicable Me* (Illumination, 2010).

In total, 87 participants watched *The Present* and 83 watched *Despicable Me*, with 66 participants overlapping, yielding 104 unique individuals (61 males, 43 females; age range = 5.0–21.9 years, mean = 12.0 ± 4.1). Additional inclusion and exclusion details are provided in Di et al. (2023).

MRI data were collected across two imaging centers—Rutgers University Brain Imaging Center (RUBIC) using a 3T Siemens Trio scanner, and the Citigroup Biomedical Imaging Center (CBIC) using 3T Siemens Prisma scanners. Scanning protocols were harmonized across sites. Functional MRI parameters were: TR = 800 ms, TE = 30 ms, flip angle = 31°, voxel size = 2.4 × 2.4 × 2.4 mm^3, and multiband acceleration factor = 6. Each run contained 250 volumes for *The Present* and 750 volumes for *Despicable Me*. High-resolution T1-weighted anatomical scans were acquired using either Human Connectome Project (HCP) or Adolescent Brain Cognitive Development (ABCD) protocols and were used solely for preprocessing. Additional acquisition details can be found on the HBN website and in Alexander et al. (2017).

2.1.2. Partly Cloudy Dataset

The *Partly Cloudy* dataset was accessed from OpenNeuro (https://openneuro.org/datasets/ds000228; Richardson et al., 2018). It included two groups: adults (n = 29; 17 females, 12 males; age range = 18–39 years, mean = 24.6 ± 5.3) and children (n = 53; 28 females, 25 males; age range = 3.5–12.3 years, mean = 7.0 ± 2.5).

Participants watched a silent version of the Pixar short film *Partly Cloudy* (5.6 minutes; https://www.pixar.com/partly-cloudy) during scanning on a 3T Siemens Tim Trio system. Younger children were scanned using one of two 32-channel custom head coils, while older children and adults used the standard Siemens 32-channel coil. Functional images were acquired using a gradient-echo EPI sequence (TR = 2000 ms, TE = 30 ms, flip angle = 90°) with 32 interleaved near-axial slices (EPI factor = 64).

Voxel sizes and slice gaps varied slightly across subgroups: (a) 3.13 mm isotropic, no gap; (b) 3.13 mm isotropic, 10% gap; (c) 3 mm isotropic, 20% gap; and (d) 3 mm isotropic, 10% gap. All

functional images were resampled to 3 mm isotropic voxels during preprocessing. Each run contained 168 volumes, with the first four discarded to allow for steady-state magnetization. T1-weighted anatomical images were collected in 176 interleaved sagittal slices (1 mm isotropic; GRAPPA = 3; FOV = 256 mm). Further methodological details are provided in Richardson et al. (2018).

2.2. Eye tracking data

Although eye-tracking data were not collected during the MRI sessions, a subset of participants from the HBN project participated in a separate EEG session conducted on a different day outside the scanner (Langer et al., 2017). During this session, participants watched several videos, including *The Present*, a shorter segment of *Despicable Me*, *Fun with Fractals* (an educational video), and a trailer for *Diary of a Wimpy Kid*, while simultaneous EEG and eye-tracking data were recorded. Eye position and pupil size were measured using an infrared video-based eye tracker (iView-X Red-m, SensoMotoric Instruments GmbH) operating at sampling rates of 30, 60, or 120 Hz. In the current study, only pupil size data were analyzed. All available recordings were included, but participants were excluded if more than 10% of time points were missing, typically due to blinks or signal loss. The final sample consisted of 1,312 participants (665 males, 387 females, 260 sex not reported), ranging in age from 5 to 21 years (mean = 9.94, SD = 3.24).

2.3. Subjective Rating Data Collection

Subjective rating data were collected from a separate group of participants at the New Jersey Institute of Technology (NJIT) to capture continuous, moment-to-moment fluctuations in perceived mental states of movie characters. A total of 34 participants (age range: 18–27 years, mean ± SD = 20.9 ± 2.3 years; 17 female) took part in the study. All participants reported normal or corrected-to-normal vision and hearing, and none had a history of neurological or psychiatric disorders. Informed consent was obtained from all participants in accordance with the Institutional Review Board (IRB) of NJIT.

Participants viewed the same three video clips used in the fMRI experiment. Each session took place in a quiet, darkened room to minimize visual and auditory distractions. Stimuli were presented on a computer monitor with synchronized sound through a speaker.

Continuous subjective ratings were acquired using CARMA software (Continuous Affect Rating and Media Annotation; (Girard, 2014) with a gaming joystick as the input device. Participants were instructed to continuously indicate the extent to which they could perceive the characters' emotions, thoughts, or intentions while watching each clip. This measure was designed to capture participants' theory-of-mind engagement—that is, the degree to which they inferred or empathically understood the internal mental states of the characters throughout the narrative. Before the experiment, participants completed a short practice session to familiarize themselves with the rating procedure.

The joystick's vertical position corresponded to a continuous numerical scale ranging from -10 (not at all) to 10 (very strongly), representing the perceived strength of mental-state understanding at each time point. Rating data were sampled at 4 Hz (every 250 ms) to provide a high-resolution temporal record of evolving subjective experience. Between clips, participants were given brief rest periods to minimize fatigue.

2.4. Video feature extraction

We extracted three sensory features from the videos: mean pixel intensity, standard deviation of pixel intensity, and pitch. For the two visual features, each video frame was first converted to grayscale, and the mean and standard deviation of pixel intensity across the two-dimensional matrix were computed. These yielded two time series corresponding to frame luminance and frame contrast, respectively. For the auditory feature, audio signals from all channels were averaged, and pitch was computed using a 100-ms window with no overlap, resulting in a 10 Hz pitch time series. Pitch was extracted only for *The Present* and *Despicable Me*, as the *Partly Cloudy* video contains no sound. Figure 2 shows the time series of mean and standard deviations of pixel intensity and pitch for the video *The Present*.

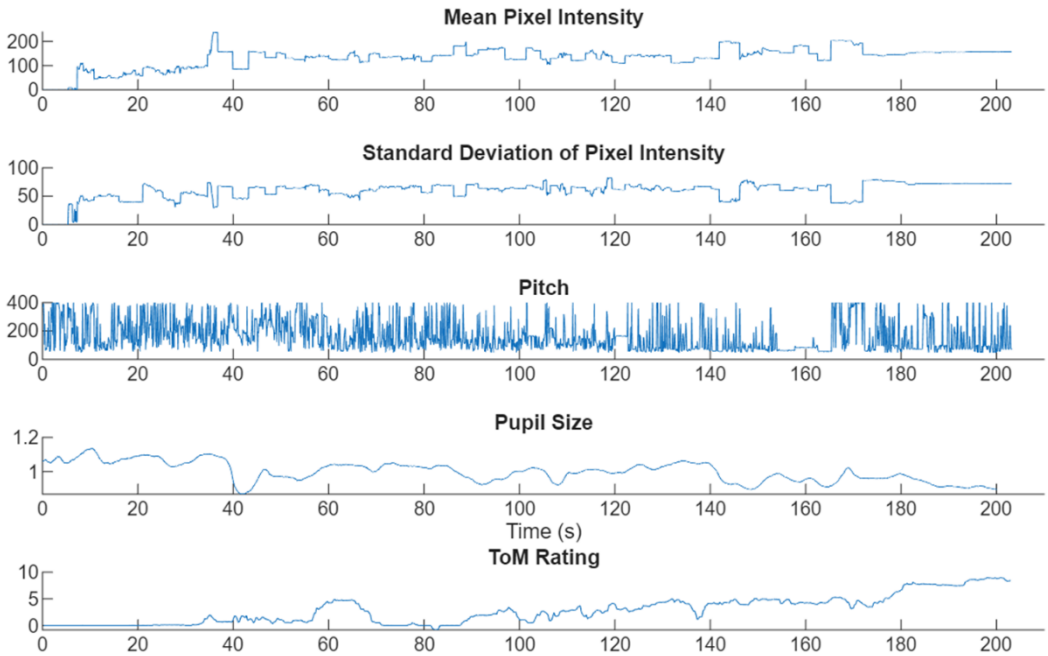

Figure 2 Example time series of different features associated with the video *The Present*.

2.5. Pupil Size Data Processing

Pupil size data for each participant were first resampled to a uniform sampling rate of 30 Hz. Missing values (NaNs) were interpolated linearly, and blink periods—identified as time points with zero values—were also replaced using linear interpolation. Participants with more than 10% of time points classified as blinks were excluded from further analyses. To reduce inter-individual variability in absolute pupil size, each participant's pupil time series was normalized by dividing by its mean.

To evaluate the consistency of pupil size fluctuations across participants, we computed leave-one-out inter-individual correlations. For each participant, their pupil size time series was correlated with the mean time series of all other participants, resulting in one correlation coefficient per individual. The statistical significance of the group-level mean correlation was assessed using a circular time-shift randomization procedure with 1,000 iterations to generate a null distribution (Kauppi et al., 2010). Finally, the median pupil size time series across participants was extracted and used as the representative group-level signal for subsequent analyses (Figure 2).

2.6. Rating Data Processing

The subjective rating data were analyzed using the same leave-one-out inter-individual correlation approach described above. For each video clip, the median rating time series across participants was computed and used as the group-level representative signal for further analyses (Figure 2 shows an example).

2.7. MRI Data Processing

Preprocessing of the fMRI data was performed separately for the HBN and *Partly Cloudy* datasets, following procedures described in previous studies. Here, we provide a brief overview. Structural MRI images for each participant were first segmented and normalized to the standard Montreal Neurological Institute (MNI) space. For *The Present* and *Despicable Me* datasets, the first 10 functional volumes were discarded, resulting in 240 and 740 volumes, respectively. In contrast, all functional volumes were retained for the Partly Cloudy dataset. The remaining functional images were then realigned to the first retained image, coregistered to the participant's structural image, normalized to MNI space using the deformation fields obtained during segmentation, and spatially smoothed with an 8 mm full-width-at-half-maximum (FWHM) Gaussian kernel.

Finally, a general linear model (GLM) was constructed to regress out head motion effects using Friston's 24-parameter motion model, which also incorporated an implicit high-pass filter with a cutoff frequency of 1/128 Hz. The resulting residual time series were used in all subsequent analyses.

We adopted a region-of-interest (ROI)–based approach to achieve a balance between spatial specificity and temporal flexibility. Cortical regions were defined using the Schaefer 100-region parcellation (Schaefer et al., 2018), and subcortical regions were defined using the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). Fourteen bilateral subcortical nuclei were included: hippocampus, parahippocampus, amygdala, caudate, putamen, pallidum, and thalamus. The Schaefer parcellation was divided into left (#1–50) and right (#51–100) hemispheres and further organized into

seven large-scale functional networks: visual, somatomotor, dorsal attention, salience/ventral attention, limbic, control, and default mode.

In total, 114 ROIs were analyzed. For each preprocessed fMRI run, the mean time series was extracted from each ROI, resulting in data matrices of size 240 × 114 for *The Present*, 740 × 114 for *Despicable Me*, and 168 × 114 for *Partly Cloudy* for each participant.

2.8. Cross-correlation Analysis

Each feature time series from the three movie clips was first convolved with the canonical hemodynamic response function (HRF) implemented in SPM12 (using the default parameters: peak delay = 6 s, undershoot delay = 16 s, and ratio = 6). The convolved feature time series were then resampled to match the temporal resolution of the corresponding fMRI data (TR = 0.8 s for *The Present* and *Despicable Me*; TR = 2.0 s for *Partly Cloudy*).

Subsequently, cross-correlation analyses were performed between the HRF-convolved feature time series and the mean fMRI time series from each ROI for each participant. Correlations were computed across a lag range of ±16 seconds (i.e., ±20 TRs for HBN data and ±8 TRs for *Partly Cloudy*), resulting in a lag-by-ROI correlation matrix for each participant. For visualization and group-level interpretation, these matrices were averaged across participants after Fisher's z transformation of the correlation coefficients.

In addition, a regression analysis was conducted using the HRF-convolved feature time series as predictors in a general linear model (GLM) framework. Group-level statistical significance of the resulting parameter estimates was assessed using one-sample t-tests across participants, which corresponds to testing the cross-correlation at zero lag. Multiple comparisons across the 114 ROIs were controlled using the false discovery rate (FDR) at $p < 0.05$.

To further examine temporal dependencies beyond the canonical HRF assumption, we also repeated the cross-correlation and zero-lag t-tests using the *raw* (unconvolved) time series of pupil size and theory-of-mind (ToM) ratings. Finally, we performed a simple cross-correlation analysis between

frame luminance and the group-level median pupil size time series for *The Present* video clip to explore direct visual–physiological coupling.

2.9. Finite Impulse Response Deconvolution

We performed finite impulse response (FIR) deconvolution to estimate the system response function between an input signal ($x(t)$) and an output signal ($y(t)$). Specifically, we first examined the pupillary system response using frame luminance as the input ($x(t)$) and pupil size as the output ($y(t)$). Subsequently, we applied the same approach to estimate the HRFs in different brain regions using the various video-related features described above as input signals and fMRI time series from each ROI as outputs.

In general, the output signal can be modeled as the convolution of the input signal with an unknown system response function:

$$y(t) = h(t) \otimes x(t) + \varepsilon(t)$$

where ($h(t)$) denotes the system's impulse response function, $\otimes$ represents the convolution operator, and ($\varepsilon(t)$) is the residual noise term. The goal of FIR deconvolution is to estimate ($h(t)$) directly from the observed time series.

We modeled ($h(t)$) as a finite-duration impulse response with a window of 20 seconds (equivalent to 25 TRs for *The Present* and *Despicable Me*, and 10 TRs for *Partly Cloudy*). The deconvolution was implemented using linear regression, where the input signal was expanded into a time-lagged design matrix containing shifted versions of ($x(t)$) at each time lag. The estimated regression coefficients correspond to the discrete samples of ($h(t)$). To improve numerical stability and reduce overfitting, we first applied an autoregressive AR(1) pre-whitening step, followed by second-difference regularization with a penalty term ($\lambda = 10$). Results were qualitatively consistent across robust to a range of λ values.

For the pupil–luminance analysis, FIR deconvolution was performed on the median pupil size time series. Both the pupil and luminance signals were z-scored before analysis to remove baseline offsets and normalize amplitudes.

For the fMRI data, FIR analysis was conducted separately for each participant and ROI. This procedure yielded an FIR kernel characterizing the temporal dynamics of the response to each input feature. Group-level response functions were obtained by averaging the individual FIR estimates after temporal alignment. Prior to deconvolution, input features were resampled to match the fMRI sampling rate, and both input and output signals were z-scored. The resulting FIR functions were visualized to examine the temporal profiles and peak latencies of the responses, allowing comparison of pupil and neural dynamics across features and brain regions.

3. Results

3.1. Sensory features

We first examined the correspondence between HRF-convolved sensory feature time series and regional BOLD activity using cross-correlation analyses across 114 regions of interest (ROIs). For each ROI, cross-correlation functions were computed for individual participants, averaged across subjects, and concatenated across regions to form a temporal–spatial correlation map.

Figure 3 illustrates the cross-correlation pattern for the HRF-convolved mean pixel intensity (frame luminance). The vertical axis represents temporal lags between the feature and the BOLD signals, with the center line (y = 0) indicating zero delay. Prominent peaks near zero lag suggest that convolution with the canonical HRF effectively captured the hemodynamic delay and aligned the two signals. As expected, strong correlations were consistently observed in visual cortical regions across all three videos (ROIs #1–9 in the left hemisphere and #51–58 in the right hemisphere), reflecting the tight coupling between luminance fluctuations and visual processing. In *Partly Cloudy*, the correlation peaks appeared slightly earlier, yet the zero-lag correlation remained positive, indicating that simple correlation still captured luminance-driven responses. The spatial overlap of significant zero-lag correlations across

videos (Figure 3D) further supports this observation while highlighting variability in both temporal dynamics and spatial distribution of luminance-related brain responses across stimuli.

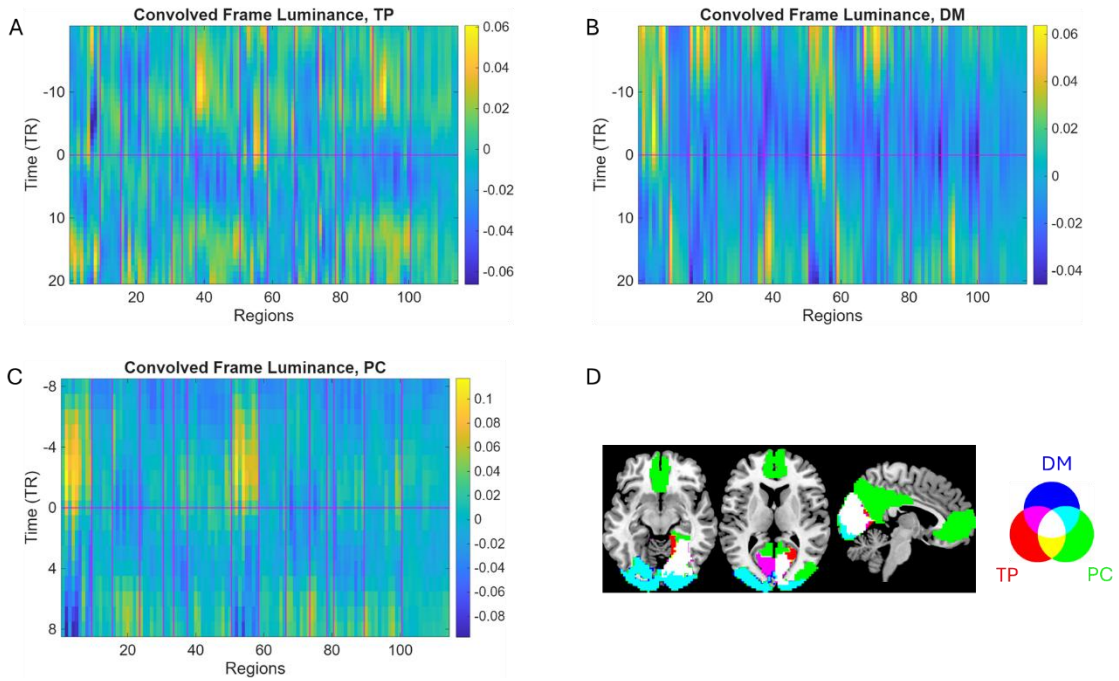

Figure 3. Cross-correlations between the hemodynamic response function (HRF)–convolved mean pixel intensity (frame luminance) of the videos and functional MRI (fMRI) time series across 114 brain regions for the three movie clips: *The Present* (A), *Despicable Me* (B), and *Partly Cloudy* (C). Vertical lines in each matrix delineate Yeo's seven-network parcellations for the left and right hemispheres, as well as subcortical regions. Panel (D) shows brain regions exhibiting significant correlations at zero lag across the three video clips, corrected for multiple comparisons using a false discovery rate (FDR) threshold of $p < 0.05$.

The other two sensory features—frame contrast (measured as the standard deviation of pixel intensity) and auditory pitch—showed comparable temporal profiles. Both features produced peak correlations near zero lag within modality-specific sensory cortices (visual and auditory, respectively). In some videos, correlations extended modestly into higher-order association areas, suggesting potential contributions from contextual or integrative processing beyond primary sensory regions (Figures 4 and 5).

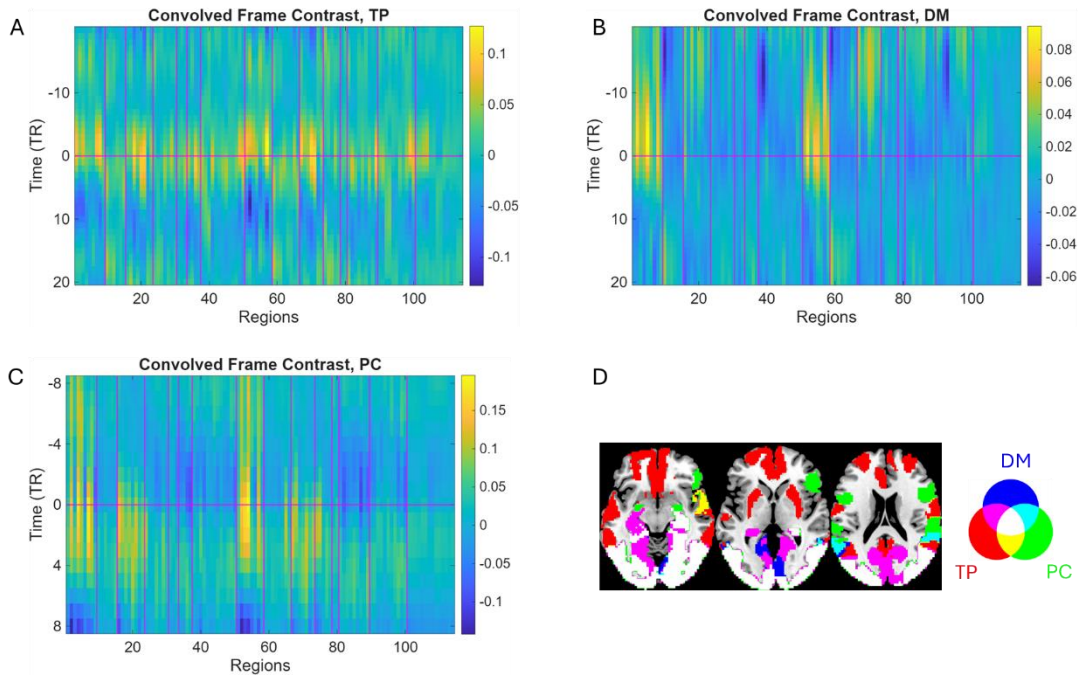

Figure 4. Cross-correlations between the hemodynamic response function (HRF)–convolved mean pixel standard deviation (frame contrast) of the videos and functional MRI (fMRI) time series across 114 brain regions for the three movie clips: The Present (A), Despicable Me (B), and Partly Cloudy (C). Vertical lines in each matrix delineate Yeo's seven-network parcellations for the left and right hemispheres, as well as subcortical regions. Panel (D) shows brain regions exhibiting significant correlations at zero lag across the three video clips, corrected for multiple comparisons using a false discovery rate (FDR) threshold of $p < 0.05$.

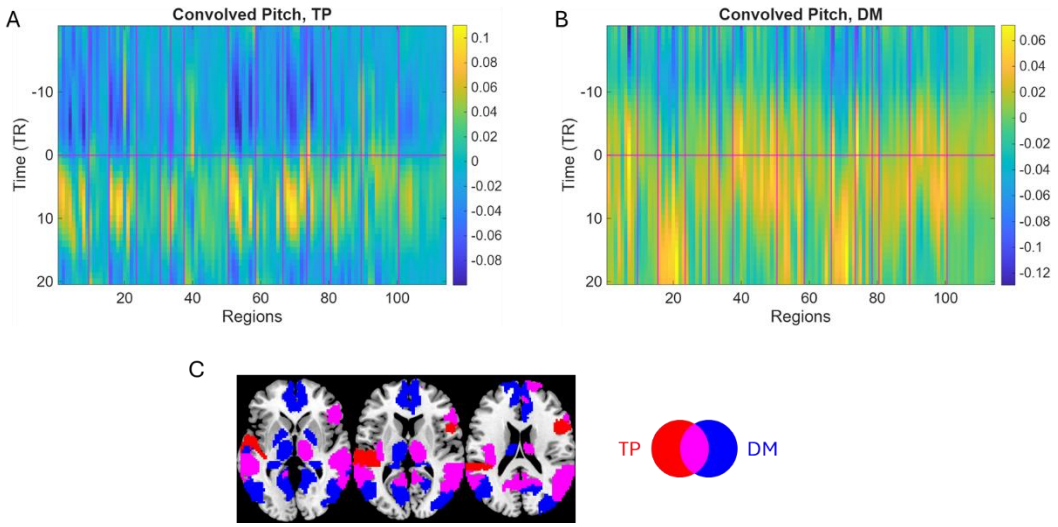

Figure 5. Cross-correlations between the hemodynamic response function (HRF)–convolved auditory pitch of the videos and functional MRI (fMRI) time series across 114 brain regions for the three movie clips: The Present (A) and Despicable Me (B). The video Partly Cloudy doesn't have sound, so was not analyzed for pitch. Vertical lines in each matrix delineate Yeo's seven-network parcellations for the left and right hemispheres, as well as subcortical regions. Panel (C) shows brain regions exhibiting significant correlations at zero lag across the three video clips, corrected for multiple comparisons using a false discovery rate (FDR) threshold of $p < 0.05$.

3.2. Pupil Sizes

The inter-individual correlation of pupil size time series for the video *The Present* was *0.62* ($p < 0.001$) across 1,312 participants, indicating a high degree of consistency in pupil dynamics. This strong concordance supports the use of the median pupil size time series as a representative measure for subsequent fMRI analyses.

We first examined its temporal relationship between frame luminance and pupil size. Figure 6A illustrates the median pupil size during the video *The Present* alongside the corresponding frame luminance. While the two time series showed a clear inverse relationship, a noticeable delay was

observed between changes in frame luminance and subsequent pupil constriction. Cross-correlation analysis confirmed this pattern, revealing a distinct negative peak around a 5-second lag (Figure 6B), indicating that pupil size decreases approximately 5 seconds after luminance increases. This temporal relationship was consistently replicated across three additional videos from the HBN-EEG dataset, which did not include fMRI data.

We further estimated the pupil response function of the median pupil time series using a finite impulse response (FIR) model (Figure 6C). For comparison, a canonical hemodynamic response function (HRF) is shown alongside. Although the two functions differed in overall shape, their peak timings were remarkably similar. Given that FIR provides a relatively unconstrained estimate of system dynamics, a smoothed version might more closely resemble the canonical HRF. Nonetheless, these findings raise an important methodological consideration: if the pupil response already exhibits a delay comparable to the HRF, applying an additional HRF convolution when mapping brain–pupil relationships may be unnecessary and could potentially distort temporal alignment.

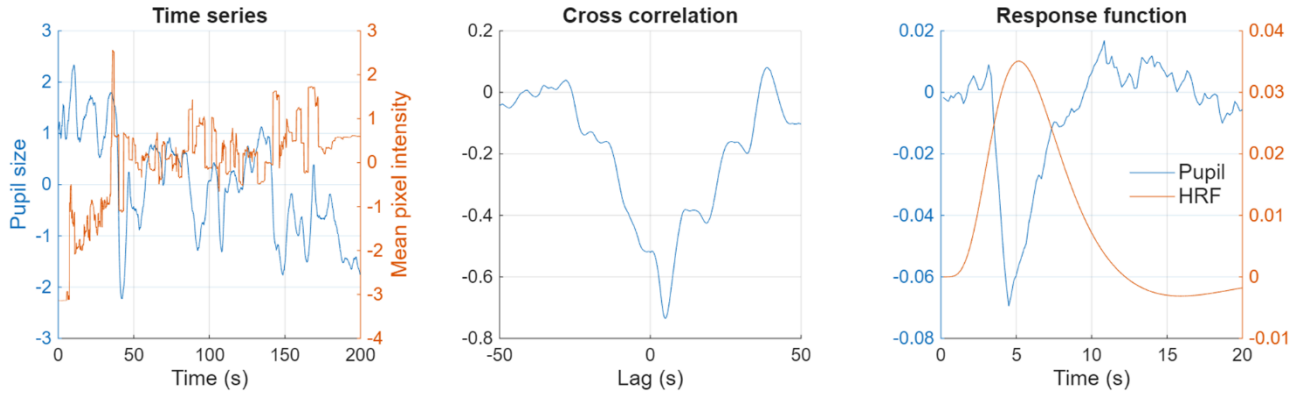

Figure 6. *(A)* Time series of median pupil size and frame luminance (mean pixel intensity) during the movie *The Present*. *(B)* Cross-correlation between the two time series, showing a negative peak indicating delayed pupil constriction following luminance changes. *(C)* Estimated system response function of pupil size using FIR deconvolution with frame luminance as the input. The canonical hemodynamic response function (HRF) is shown for comparison.

We next conducted cross-correlation analyses between BOLD signals and both the HRF-convolved and raw pupil size time series. We first focused on the visual regions that had shown strong correlations with convolved frame luminance (Figure 3). Theoretically, pupil size should be negatively correlated with activity in these regions, as increased luminance leads to pupil constriction and heightened visual cortical activation. However, when the pupil size time series was convolved with the canonical HRF, no clear negative correlation was observed at zero lag (Figure 7A). In contrast, using the raw pupil size time series revealed robust negative correlations with visual cortical activity at lag zero (Figure 7B).

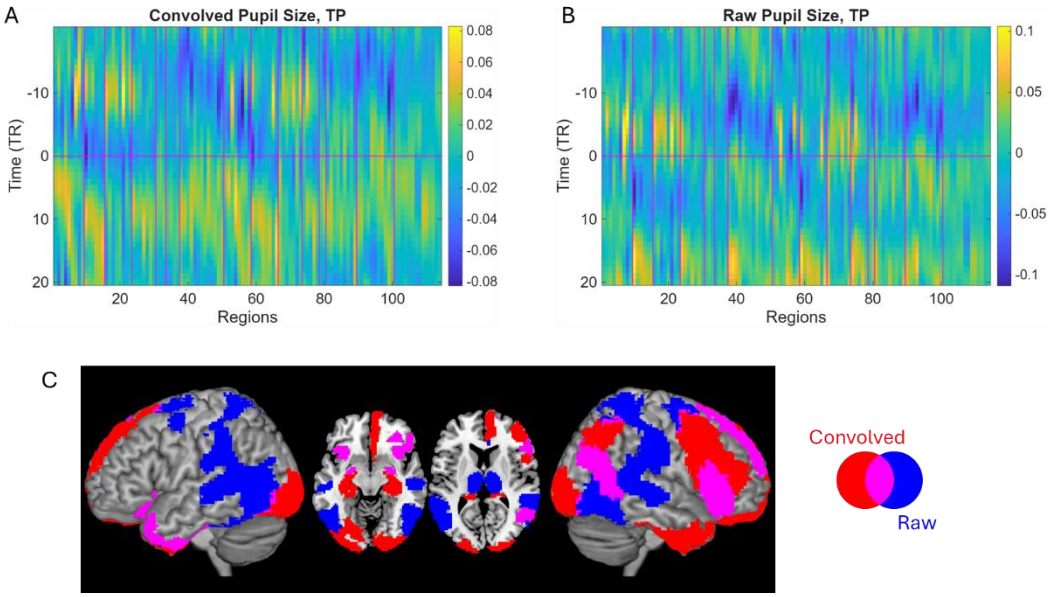

Figure 7. Cross-correlations between the hemodynamic response function (HRF)–convolved (A) and raw (B) pupil size and functional MRI (fMRI) time series across 114 brain regions for the movie *The Present*. Vertical lines in each matrix delineate Yeo's seven-network parcellations for the left and right hemispheres, as well as subcortical regions. Panel (C) shows brain regions exhibiting significant correlations at zero lag for the convolved and raw pupil size time series, corrected for multiple comparisons using a false discovery rate (FDR) threshold of $p < 0.05$.

In addition, the spatial distribution of correlated regions differed substantially between the two approaches. Overlapping correlations were observed in the bilateral insula and right temporoparietal

junction, but the raw pupil size time series revealed broader and more distinct correlations across lateral parietal and temporal cortices as well as the thalamus. These findings suggest that HRF convolution may distort the temporal alignment between pupil dynamics and BOLD responses, whereas using the raw pupil time series preserves more physiologically meaningful relationships.

3.3. Subjective Rating of Theory of Mind

The inter-individual correlations of ToM rating time series were 0.58 for *The Present*, 0.41 for *Despicable Me*, and 0.59 (all ps < 0.001) for *Partly Cloudy*, based on circular time-shift randomization. These medium-to-strong correlations indicate a high consistency in subjective ratings, supporting the use of the median rating time series as a representative measure for subsequent fMRI analyses.

We next analyzed the subjective rating data. Using the median rating time series convolved with the canonical HRF, we computed cross-correlations with fMRI activity (Figure 8). For *The Present* (TP), stronger correlations were observed in multiple regions—primarily within visual and dorsal attention networks—with notable delays. In *Despicable Me* (DM), correlations were generally weaker, though delayed positive correlations were still evident in the dorsal attention network (ROIs 16–23 in the left hemisphere and 67–73 in the right hemisphere). Interestingly, two prominent peaks at advanced (negative) lags appeared around ROIs 40 and 90, corresponding to auditory regions—an implausible finding given their limited involvement in Theory of Mind processes. For *Partly Cloudy* (PC), the HRF-convolved time series aligned more effectively with the fMRI signals, showing peak correlations near zero lag. Finally, zero-lag correlation maps across the three videos (Figure 8D) revealed only one overlapping region (ROI #7) in the left occipito-parietal region, indicating limited spatial consistency in Theory of Mind–related activity across stimuli.

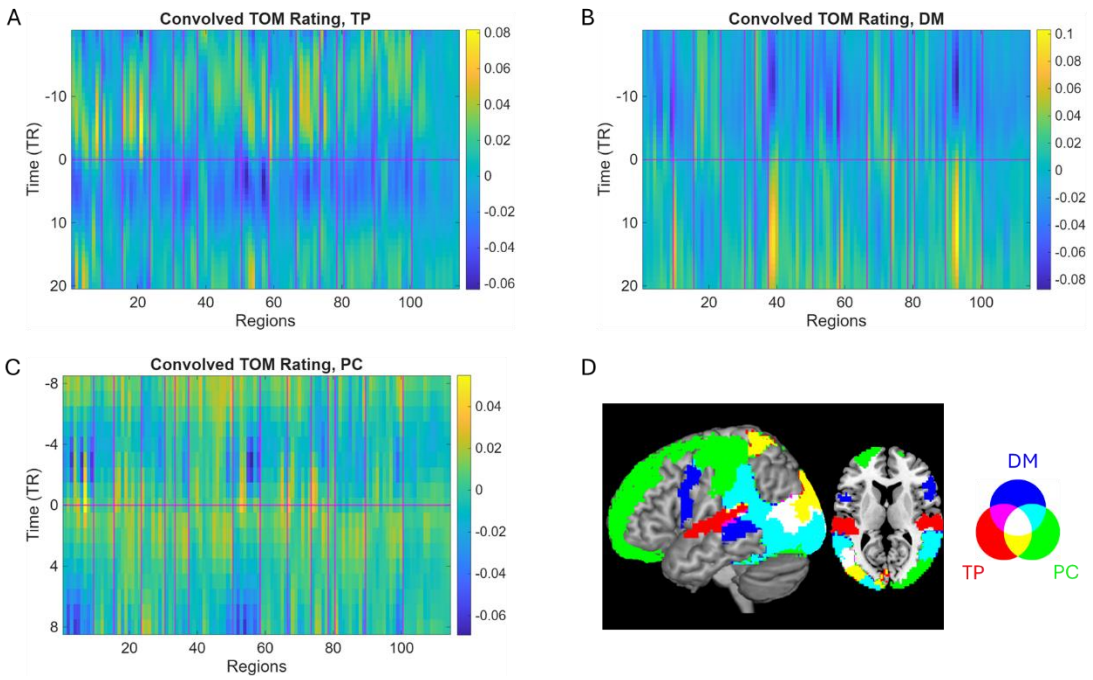

Figure 8. Cross-correlations between the hemodynamic response function (HRF)–convolved theory of mind rating and functional MRI (fMRI) time series across 114 brain regions for the three movie clips: *The Present* (A), *Despicable Me* (B), and *Partly Cloudy* (C). Vertical lines in each matrix delineate Yeo's seven-network parcellations for the left and right hemispheres, as well as subcortical regions. Panel (D) shows brain regions exhibiting significant correlations at zero lag across the three video clips, corrected for multiple comparisons using a false discovery rate (FDR) threshold of $p < 0.05$.

Next, we performed cross-correlation analyses using the raw (unconvolved) ToM rating time series. For *The Present* (TP), clear peaks around zero lag were observed in visual and dorsal attention regions. A similar pattern emerged for *Despicable Me* (DM), though the peaks were broader and less pronounced. For *Partly Cloudy* (PC), several regions—such as ROIs 7 and 57—also showed peak correlations near zero lag, while others exhibited slightly advanced peaks, resulting in weaker zero-lag correlations. Finally, zero-lag correlation maps across the three videos (Figure 9D) revealed overlapping regions in the left occipito-temporal junction (ROI #7 and #16), suggesting larger overlapping than using convolved ToM rating time series.

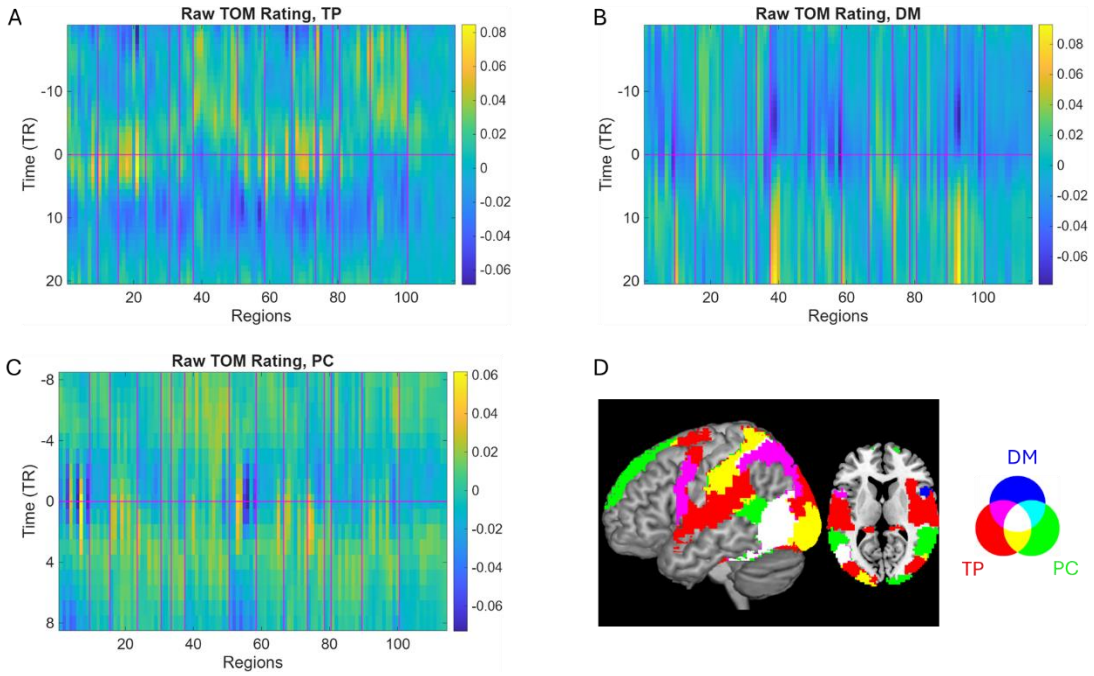

Figure 9 Cross-correlations between the raw theory of mind rating and functional MRI (fMRI) time series across 114 brain regions for the three movie clips: *The Present* (A), *Despicable Me* (B), and *Partly Cloudy* (C). Vertical lines in each matrix delineate Yeo's seven-network parcellations for the left and right hemispheres, as well as subcortical regions. Panel (D) shows brain regions exhibiting significant correlations at zero lag across the three video clips, corrected for multiple comparisons using a false discovery rate (FDR) threshold of $p < 0.05$.

3.4. Finite Impulse Response Analysis

Lastly, we performed finite impulse response (FIR) deconvolution for the five features across the three videos, focusing on representative brain regions associated with each feature (Figure 10). For frame luminance, we selected a medial occipital region that showed consistent correlations across all three videos. The estimated response function exhibited a peak around 5 seconds, broadly consistent with the canonical HRF. However, substantial variability was observed across videos, particularly a pronounced dip around 13 seconds for *Partly Cloudy*. For frame contrast, we examined a lateral occipital region with reliable contrast-related activity. The estimated responses showed peaks at approximately 3–4 seconds for

The Present and *Despicable Me*, whereas the response for *Partly Cloudy* was noticeably delayed. Finally, for auditory pitch, we analyzed an auditory cortical region, where the estimated responses for *The Present* and *Despicable Me* closely resembled the canonical HRF. The response seemed slower to reach its peak for The Present compared to Despicable Me.

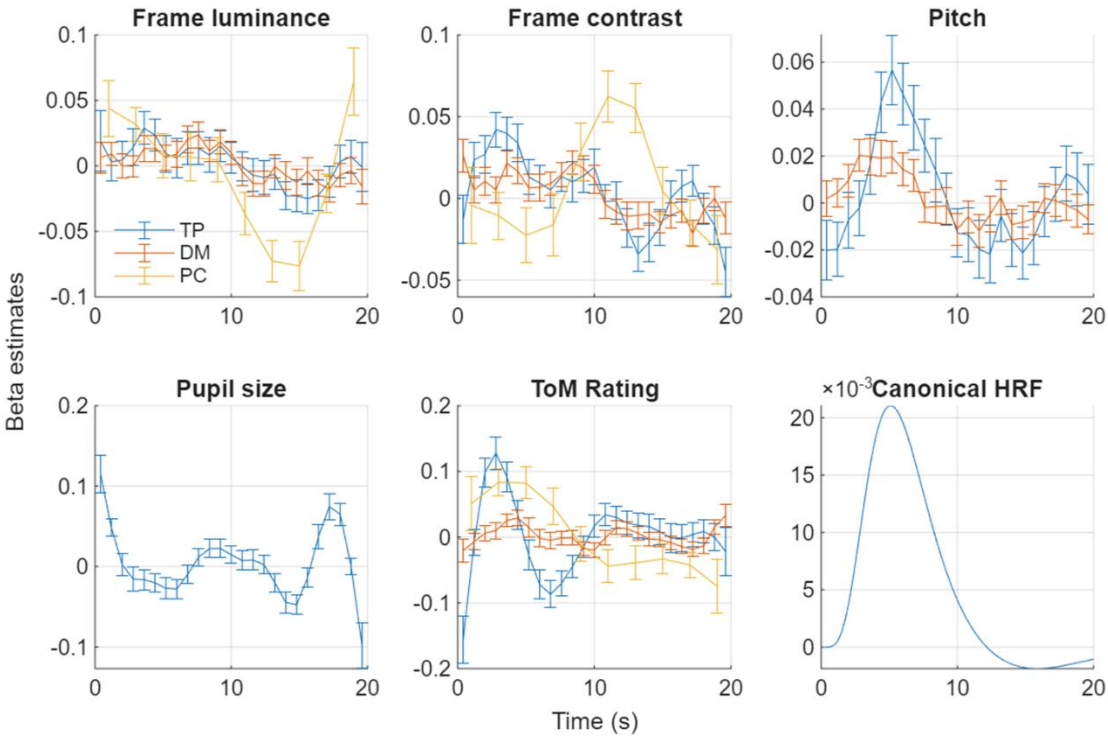

Figure 10. Finite impulse response (FIR) deconvolution results for the five features across the three videos, shown for one representative brain region per feature. Error bars indicate 95% confidence intervals across participants. A canonical hemodynamic response function (HRF) from SPM is displayed in the bottom right panel for reference.

We selected a lateral occipital region to illustrate the response function for pupil size. The estimated response function peaked at the first time point, consistent with the cross-correlation results, suggesting that pupil size fluctuations are nearly synchronous with corresponding BOLD signal changes in this region.

Lastly, we examined a left lateral occipital region to display the estimated response functions for theory-of-mind (ToM) ratings across the three videos. The response function for *The Present* rose more rapidly than the canonical HRF, while those for *Despicable Me* and *Partly Cloudy* were also somewhat faster and broader. This pattern suggests that the true underlying responses to ToM-related fluctuations may exhibit shorter latency and greater temporal dispersion compared with the canonical HRF. Although convolving ToM ratings with the canonical HRF may still capture relevant activations, using a more flexible or data-driven HRF could provide a closer approximation of the underlying neural dynamics.

4. Discussion

By performing cross-correlation analysis and FIR deconvolution, the current results can be summarized as follows:

1. For visual and acoustic features, canonical HRF could effectively model the hemodynamic response delays.

2. Pupil size response to frame luminance changes had a response delay that is comparable to canonical HRF. Therefore, using raw pupil size time series could capture more meaningful brain activity than using the HRF convolved version.

3. For the subjective-rating data, the delay might vary across videos. But in general, using raw time series, rather than the HRF convolved version, showed more consistent correlations in brain regions across videos.

4.1. Sensory features

The present study demonstrated that for low-level sensory features—specifically frame luminance, frame contrast, and acoustic pitch—the canonical hemodynamic response function (HRF) implemented in standard fMRI analysis tools (e.g., SPM) effectively captured the temporal delay between stimulus dynamics and BOLD responses. Cross-correlation analyses revealed that the strongest correlations between these features and regional BOLD signals occurred at lags consistent with the canonical HRF

peak (approximately 4–6 seconds). Similarly, finite impulse response (FIR) deconvolution yielded impulse response functions that closely mirrored the canonical HRF, exhibiting the characteristic rise, peak, and undershoot phases. These results indicate that the temporal characteristics of the BOLD response to continuously varying visual and auditory inputs are largely consistent with those reported in conventional event-related paradigms, even under naturalistic viewing conditions.

This finding aligns with classic event-related fMRI studies showing that BOLD responses in early sensory regions are highly stereotyped in both shape and latency (Boynton et al., 1996; Dale and Buckner, 1997; Glover, 1999; Handwerker et al., 2004). It also resonates with more recent research using naturalistic stimuli, which has shown that BOLD activity in visual and auditory cortices remains tightly coupled to sensory fluctuations over time (Bartels and Zeki, 2004; Hasson et al., 2004; Nishimoto et al., 2011). Together, these studies support the notion that the canonical HRF offers a robust approximation of average neurovascular coupling in primary sensory areas across diverse experimental contexts.

Nonetheless, subtle deviations in the FIR-estimated responses across cortical regions were evident, suggesting regional variability in HRF dynamics. Such differences are consistent with reports that HRF latency and shape vary across the cortex due to differences in vascular structure, baseline cerebral blood flow, and local neurovascular coupling (Aguirre et al., 1998; Handwerker et al., 2004; Sonkusare et al., 2019; Taylor et al., 2018). Moreover, studies using high-resolution fMRI and model-free approaches have demonstrated that higher-order regions often exhibit delayed or temporally dispersed responses compared to primary sensory cortices (Lewis et al., 2018), further emphasizing the need for flexible models.

4.2. Pupil size

Previous fMRI studies have often modeled pupil diameter as a physiological regressor by convolving it with a canonical hemodynamic response function (HRF) before correlating it with BOLD signals (Schneider et al., 2016; Yellin et al., 2015). This approach assumes that pupil size reflects instantaneous neural or arousal-related processes that are temporally blurred by the hemodynamic response. However,

553 our cross-correlation analyses indicate that this assumption may not hold uniformly across brain regions
554 or datasets.

555 When comparing correlations between BOLD activity and both HRF-convolved and raw pupil
556 size time series, clear differences emerged. In visual cortical areas, where BOLD activity positively tracks
557 luminance, pupil size is expected to show an inverse relationship due to its light reflex (Joshi and Gold,
558 2020; Mathot, 2018). Consistent with this prediction, negative correlations were observed between raw
559 pupil size and visual cortex BOLD signals near zero lag. In contrast, when using the HRF-convolved
560 pupil signal, these expected negative correlations disappeared, suggesting a temporal misalignment
561 between modeled and actual pupil–BOLD dynamics. Furthermore, the spatial extent of correlated regions
562 differed markedly between the two approaches. While both revealed significant correlations in regions
563 such as the insula and temporoparietal junction, the raw pupil signal additionally showed broader
564 associations encompassing the lateral parietal and temporal cortices, as well as the thalamus—areas
565 known to participate in arousal, attention, and salience processing (Menon and Uddin, 2010; Sadaghiani
566 and D'Esposito, 2015; Shine et al., 2019).

567 These findings suggest that applying a canonical HRF convolution to pupil data may distort or
568 obscure the true temporal relationship between pupil fluctuations and neural activity. Nevertheless, this
569 does not invalidate prior work employing HRF-convolved pupil regressors. Variability in eye-tracking
570 systems, temporal resolution, and luminance control can influence the temporal characteristics of pupil
571 responses (Hoeks and Levelt, 1993; McDougal and Gamlin, 2015), and in some cases—particularly with
572 lower sampling rates—HRF convolution may remain a pragmatic approximation. However, our results
573 underscore the importance of empirically testing temporal alignment between pupil and BOLD signals
574 rather than assuming a fixed hemodynamic delay. Cross-correlation analyses, as illustrated here, offer a
575 straightforward diagnostic approach for validating these relationships and optimizing analytical strategies.

576 From a neurophysiological perspective, pupil diameter is a reliable peripheral marker of central
577 arousal and neuromodulatory activity, particularly reflecting the influence of the locus coeruleus–
578 norepinephrine (LC–NE) system (Aston-Jones and Cohen, 2005; Joshi et al., 2016; Murphy et al., 2014).

The widespread correlations observed between raw pupil fluctuations and BOLD activity in the insula, temporoparietal junction, and thalamus are consistent with the distributed impact of the LC–NE network on cortical and subcortical targets (Breton-Provencher and Sur, 2019; Shine et al., 2019). These regions are frequently implicated in arousal regulation, interoceptive awareness, and adaptive cognitive control. Thus, the coupling observed here may reflect dynamic coordination between neuromodulatory arousal signals and cortical processing of salient environmental events during naturalistic viewing.

Overall, these results emphasize the need for careful temporal modeling of pupil–BOLD coupling. Because pupil responses themselves exhibit intrinsic latency relative to underlying neural events (de Gee et al., 2017; Hoeks and Levelt, 1993), applying an additional HRF convolution can introduce systematic misalignment, reducing sensitivity to true neural–autonomic interactions. Future studies should therefore consider data-driven approaches—such as empirical lag estimation or joint modeling frameworks (Cohen et al., 2016; Di and Biswal, 2019)—to better capture the temporal dynamics linking arousal-related physiology and brain activity, both in naturalistic and controlled paradigms.

4.3. ToM Rating

In this study, an independent group of participants provided continuous ratings reflecting their perceived understanding of the characters' thoughts, feelings, and intentions while viewing short movie clips. These ratings are intended to index moment-to-moment engagement of Theory of Mind (ToM)—the ability to attribute mental states to others (Frith and Frith, 2006; Schurz et al., 2021). Unlike low-level sensory or physiological features, however, the temporal coupling between such introspective, socially grounded judgments and underlying neural activity is far less constrained and may vary with narrative structure, attention, and individual interpretation (Jacoby et al., 2016; Richardson et al., 2018).

Cross-correlation analyses revealed that the temporal alignment between ToM ratings and BOLD signals differed across movies and depended on whether the rating time series was convolved with a canonical HRF. When the HRF-convolved ratings were used, correlations tended to peak at delayed lags,

605 particularly in visual cortical regions for *The Present*. In contrast, using the raw (unconvolved) ToM
606 ratings yielded more consistent correlations across all three videos, with peaks closer to zero lag in
607 regions typically implicated in social and perceptual processing—such as the left occipito-temporal
608 junction, temporoparietal cortex, and posterior superior temporal sulcus (Schurz et al., 2014; Wheatley et
609 al., 2007). Advanced (negative-lag) peaks observed in auditory regions suggested possible temporal
610 misalignment, as neural activity cannot precede the reported subjective perception.

611 These findings suggest that convolving subjective rating time series with a canonical HRF does
612 not necessarily improve their temporal correspondence with neural signals and may, in some cases, distort
613 the timing relationship. FIR modeling further supported this interpretation: the estimated hemodynamic
614 responses to the ToM ratings were often faster or temporally broader than predicted by the canonical
615 HRF. Continuous introspective ratings inherently integrate information over several seconds due to
616 cognitive evaluation and motor execution (Nummenmaa et al., 2012; Zahn et al., 2009). Applying an
617 additional HRF convolution may thus compound temporal smoothing and shift the regressor further from
618 the true neural dynamics underlying ToM engagement.

619 Nevertheless, the precise timing of subjective reports remains uncertain, as behavioral latencies—
620 including decision-making, motor response delays, and individual differences in introspective pace—
621 introduce variability. Moreover, ToM engagement likely reflects a temporally extended inferential
622 process rather than discrete events, making a single fixed delay implausible (Jacoby et al., 2016; Spunt
623 and Adolphs, 2014). These complexities highlight the challenges of mapping continuous, high-level
624 social cognition onto temporally sluggish BOLD signals.

625 Our results converge with recent efforts advocating explicit examination of temporal relationships
626 between behavioral regressors and neural responses in naturalistic paradigms (Di and Biswal, 2022;
627 Lahnakoski et al., 2012b). Rather than assuming a canonical HRF, data-driven approaches — including
628 voxel-wise temporal-shift estimation (Liao et al., 2002), cross-correlation mapping and lag correction
629 (Siegel et al., 2016), FIR deconvolution, or HRF reconstruction methods (Herrnberger, 2022) — can be
630 used to estimate and correct hemodynamic delays in fMRI. Future work could integrate synchronized

physiological or behavioral measures (e.g., eye-tracking, facial EMG, or reaction-based time stamping) to refine lag estimation and improve interpretability.

In summary, the current results suggest that raw, unconvolved ToM ratings may better capture brain–behavior coupling during naturalistic movie viewing. However, the evidence remains preliminary, underscoring the need for careful consideration of timing assumptions when modeling continuous subjective measures in fMRI studies of social cognition.

4.4. Broader implications for hemodynamic response

The present analysis provides preliminary evidence that the HRF can be estimated directly from continuous naturalistic data using FIR modeling. Traditional HRF estimation approaches have primarily relied on event-related or block designs, where each stimulus or trial is modeled as a discrete impulse (Dale and Buckner, 1997; Glover, 1999). Although such methods have been instrumental in characterizing canonical HRF properties, they inherently simplify the temporal and cognitive structure of real-world experiences. Event-based paradigms constrain the complexity and duration of modeled processes and often assume that cognitive or affective responses occur as isolated, time-locked events rather than as continuously evolving states.

In contrast, naturalistic paradigms offer an alternative framework by leveraging time-varying sensory, behavioral, or physiological features as continuous inputs to estimate neural response functions. This approach captures more ecologically valid dynamics of neural activity, encompassing sustained attention, emotional engagement, and social inference—processes that unfold gradually rather than discretely (Hasson et al., 2004; Nastase et al., 2019). Although continuous designs face challenges such as limited variability in specific features and uneven temporal sampling across frequency ranges, they allow HRF estimation in higher-order cortical and subcortical regions that are less reliably engaged by simple, transient stimuli.

The current findings demonstrate the feasibility of using continuous naturalistic features to perform data-driven HRF deconvolution across multiple brain regions. The FIR model, which imposes

657 minimal assumptions about response shape or latency, revealed potentially diverse temporal profiles
658 across regions, consistent with prior evidence that HRF dynamics vary with vascular and neuronal
659 properties (Aguirre et al., 1998; Handwerker et al., 2004; Lindquist et al., 2009). Such flexibility is
660 valuable for uncovering region-specific or stimulus-specific variations in hemodynamic coupling that are
661 not well captured by canonical models.

662 However, this flexibility comes at the cost of increased variance and reduced reliability,
663 particularly when the number of model parameters approaches the available data length (Goutte et al.,
664 2000; Pedregosa et al., 2015). Future research should explore hybrid or regularized approaches—such as
665 ridge-regularized FIR models, basis function modeling (e.g., gamma or spline basis sets), or parametric
666 HRF estimation—to balance flexibility with stability and physiological interpretability (Lindquist et al.,
667 2009; Pedregosa et al., 2015). Comparative evaluations across these frameworks will be critical for
668 identifying optimal trade-offs between model adaptability and robustness in naturalistic contexts.

669 An additional priority is to assess the reliability and inter-individual variability of HRF estimates
670 derived from continuous data. Individual differences in vascular dynamics, neurovascular coupling, or
671 baseline physiology may systematically influence HRF latency and shape (Handwerker et al., 2004; West
672 et al., 2019). Quantifying these sources of variability could enhance both within-subject reproducibility
673 and between-subject interpretability, with potential applications in aging (Chen et al., 2024; West et al.,
674 2019), cerebrovascular disease, and neuropsychiatric populations. Robust, data-driven HRF estimation
675 could ultimately improve the interpretability of BOLD responses in naturalistic and task-free paradigms,
676 bridging the gap between experimental control and ecological validity.

677 Finally, integrating FIR-based HRF estimation with analyses of sensory, physiological, and
678 cognitive features—as demonstrated in the present study—illustrates the value of unified temporal
679 modeling across multiple domains. The observation that low-level sensory features align well with the
680 canonical HRF, whereas physiological (pupil) and introspective (ToM rating) signals exhibit more
681 complex timing relationships, suggests that hemodynamic coupling may differ systematically across
682 processing hierarchies. Data-driven HRF estimation, therefore, provides not only a methodological

advance but also a conceptual framework for understanding how distinct neural systems integrate over time. Extending this approach across diverse stimulus types and participant populations could yield a more comprehensive characterization of the temporal architecture of brain function under naturalistic conditions.

5. Conclusion

In summary, this study examined the temporal relationships between continuous sensory, physiological, and behavioral features and fMRI signals during naturalistic movie viewing. Using cross-correlation and finite impulse response (FIR) analyses, we found that low-level features such as luminance and contrast elicited rapid and consistent responses in visual regions, whereas pupil size and theory-of-mind ratings showed more variable relationships with BOLD activity in higher-order networks. These findings indicate that a single canonical hemodynamic response function may not adequately capture the diverse temporal dynamics across feature types and brain regions. Moreover, the current analysis validated the feasibility of using continuous input signals to estimate brain response functions, offering a novel approach to studying complex brain processes and higher-order cortical systems. This framework provides an alternative way to quantify hemodynamic responses, which may be particularly valuable for investigating individual differences and vascular variability across diverse populations, such as aging or clinical cohorts.

Data and Code Availability

MRI and pupil size data were obtained from publicly available sources: the Healthy Brain Network (HBN) dataset (http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/) and the OpenNeuro repository (https://openneuro.org/datasets/ds000228). The Theory of Mind (ToM) rating data have been deposited on the Open Science Framework (OSF) and will be made publicly available upon manuscript acceptance.

Author Contributions

X.D. conceived and designed the study, developed the methodology, performed the analyses, and wrote the original draft of the manuscript. G.B.H. collected and curated the data and contributed to manuscript review and editing. B.B.B. and X.D. provided resources and secured funding for the project. All authors contributed to the interpretation of results and approved the final version of the manuscript.

Funding

This study was supported by grants from (US) National Institutes of Health for Xin Di (R15MH125332) and Bharat B. Biswal (5R01MH131335 and 1R01AG085665) and by (NJ) Governor's Council for Medical Research and Treatment of Autism for Xin Di (CAUT25BRP005).

Declaration of Competing Interests

The authors declare that there is no competing interest.

Reference:

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The Variability of Human, BOLD Hemodynamic Responses. NeuroImage 8, 360–369. https://doi.org/10.1006/nimg.1998.0369

Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Andersen, J., Bronstein, B., Bui, A., Bushey, M., Butler, H., Castagna, V., Camacho, N., Chan, E., Citera, D., Clucas, J., Cohen, S., Dufek, S., Eaves, M., Fradera, B., Gardner, J., Grant-Villegas, N., Green, G., Gregory, C., Hart, E., Harris, S., Horton, M., Kahn, D., Kabotyanski, K., Karmel, B., Kelly, S.P., Kleinman, K., Koo, B., Kramer, E., Lennon, E., Lord, C., Mantello, G., Margolis, A., Merikangas, K.R., Milham, J., Minniti, G., Neuhaus, R., Levine, A., Osman, Y., Parra, L.C., Pugh, K.R., Racanello, A., Restrepo, A., Saltzman, T., Septimus, B., Tobe, R., Waltz, R., Williams, A., Yeo, A., Castellanos, F.X., Klein, A., Paus, T., Leventhal, B.L., Craddock, R.C., Koplewicz, H.S., Milham, M.P., 2017. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Scientific Data 4, 1–26. https://doi.org/10.1038/sdata.2017.181

Aston-Jones, G., Cohen, J.D., 2005. AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. Annual Review of Neuroscience 28, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bartels, A., Zeki, S., 2004. Functional brain mapping during free viewing of natural scenes. Hum Brain Mapp 21, 75–85. https://doi.org/10.1002/hbm.10153

Bartels, A., Zeki, S., Logothetis, N.K., 2008. Natural Vision Reveals Regional Specialization to Local Motion and to Contrast-Invariant, Global Flow in the Human Brain. Cerebral Cortex 18, 705–717. https://doi.org/10.1093/cercor/bhm107

Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. J. Neurosci. 16, 4207–4221. https://doi.org/10.1523/JNEUROSCI.16-13-04207.1996

Breton-Provencher, V., Sur, M., 2019. Active control of arousal by a locus coeruleus GABAergic circuit. Nat Neurosci 22, 218–228. https://doi.org/10.1038/s41593-018-0305-z

Chen, D.Y., Di, X., Yu, X., Biswal, B.B., 2024. The significance and limited influence of cerebrovascular reactivity on age and sex effects in task- and resting-state brain activity. Cerebral Cortex 34, bhad448. https://doi.org/10.1093/cercor/bhad448

Cohen, A.O., Breiner, K., Steinberg, L., Bonnie, R.J., Scott, E.S., Taylor-Thompson, K., Rudolph, M.D., Chein, J., Richeson, J.A., Heller, A.S., Silverman, M.R., Dellarco, D.V., Fair, D.A., Galván, A., Casey, B.J., 2016. When Is an Adolescent an Adult? Assessing Cognitive Control in Emotional and Nonemotional Contexts. Psychol Sci 27, 549–562. https://doi.org/10.1177/0956797615627625

Dale, A.M., Buckner, R.L., 1997. Selective averaging of rapidly presented individual trials using fMRI. Human brain mapping 5, 329–40. https://doi.org/10.1002/(SICI)1097-0193(1997)5:5%253C329::AID-HBM1%253E3.0.CO;2-5

de Gee, J.W., Colizoli, O., Kloosterman, N.A., Knapen, T., Nieuwenhuis, S., Donner, T.H., 2017. Dynamic modulation of decision biases by brainstem arousal systems. eLife 6, e23232. https://doi.org/10.7554/eLife.23232

Di, X., Biswal, B.B., 2022. Principal component analysis reveals multiple consistent responses to naturalistic stimuli in children and adults. Human Brain Mapping 43, 3332–3345. https://doi.org/10.1002/hbm.25568

Di, X., Biswal, B.B., 2019. Toward Task Connectomics: Examining Whole-Brain Task Modulated Connectivity in Different Task Domains. Cereb Cortex 29, 1572–1583. https://doi.org/10.1093/cercor/bhy055

Di, X., Xu, T., Miao, X., 2026. Chapter 14 - Future perspectives of functional connectivity: Advances in data acquisition and analysis, in: Rocca, M.A., Filippi, M. (Eds.), Functional Connectivity of the Human Brain. Academic Press, pp. 281–309. https://doi.org/10.1016/B978-0-443-19099-5.00020-9

Di, X., Xu, T., Uddin, L.Q., Biswal, B.B., 2023. Individual differences in time-varying and stationary brain connectivity during movie watching from childhood to early adulthood: Effects of age, sex, and behavioral associations. https://doi.org/10.1101/2023.01.30.526311

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-Related fMRI: Characterizing Differential Responses. NeuroImage 7, 30–40. https://doi.org/10.1006/nimg.1997.0306

Frith, C.D., Frith, U., 2006. The Neural Basis of Mentalizing. Neuron 50, 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Girard, J.M., 2014. CARMA: Software for Continuous Affect Rating and Media Annotation | Journal of Open Research Software. https://doi.org/10.5334/jors.ar

Glover, G.H., 1999. Deconvolution of Impulse Response in Event-Related BOLD fMRI1. NeuroImage 9, 416–429. https://doi.org/10.1006/nimg.1998.0419

Goutte, C., Nielsen, F.A., Hansen, K.H., 2000. Modeling the hemodynamic response in fMRI using smooth FIR filters. IEEE Transactions on Medical Imaging 19, 1188–1201. https://doi.org/10.1109/42.897811

Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21, 1639–1651. https://doi.org/10.1016/j.neuroimage.2003.11.029

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science (New York, N.Y.) 303, 1634–40. https://doi.org/10.1126/science.1089506

Herrnberger, B., 2022. Reconstruction of time-shifted hemodynamic response. Sci Rep 12, 17441. https://doi.org/10.1038/s41598-022-17601-5

Hoeks, B., Levelt, W.J.M., 1993. Pupillary dilation as a measure of attention: a quantitative system analysis. Behavior Research Methods, Instruments, & Computers 25, 16–26. https://doi.org/10.3758/BF03204445

Huth, A.G., Lee, T., Nishimoto, S., Bilenko, N.Y., Vu, A.T., Gallant, J.L., 2016. Decoding the Semantic Content of Natural Movies from Human Brain Activity. Front. Syst. Neurosci. 10. https://doi.org/10.3389/fnsys.2016.00081

Jacoby, N., Bruneau, E., Koster-Hale, J., Saxe, R., 2016. Localizing Pain Matrix and Theory of Mind networks with both verbal and non-verbal stimuli. NeuroImage 126, 39–48. https://doi.org/10.1016/j.neuroimage.2015.11.025

Joshi, S., Gold, J.I., 2020. Pupil Size as a Window on Neural Substrates of Cognition. Trends in Cognitive Sciences 24, 466–480. https://doi.org/10.1016/j.tics.2020.03.005

Joshi, S., Li, Y., Kalwani, R.M., Gold, J.I., 2016. Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. Neuron 89, 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Kauppi, J.-P., Jääskeläinen, I.P., Sams, M., Tohka, J., 2010. Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency. Front. Neuroinform. 4. https://doi.org/10.3389/fninf.2010.00005

Lahnakoski, J.M., Glerean, E., Salmi, J., Jääskeläinen, I.P., Sams, M., Hari, R., Nummenmaa, L., 2012a. Naturalistic fMRI Mapping Reveals Superior Temporal Sulcus as the Hub for the Distributed Brain Network for Social Perception. Front. Hum. Neurosci. 6. https://doi.org/10.3389/fnhum.2012.00233

Lahnakoski, J.M., Salmi, J., Jääskeläinen, I.P., Lampinen, J., Glerean, E., Tikka, P., Sams, M., 2012b. Stimulus-Related Independent Component and Voxel-Wise Analysis of Human Brain Activity during Free Viewing of a Feature Film. PLOS ONE 7, e35215. https://doi.org/10.1371/journal.pone.0035215

Langer, N., Ho, E.J., Alexander, L.M., Xu, H.Y., Jozanovic, R.K., Henin, S., Petroni, A., Cohen, S., Marcelle, E.T., Parra, L.C., Milham, M.P., Kelly, S.P., 2017. A resource for assessing information processing in the developing brain using EEG and eye tracking. Sci Data 4, 170040. https://doi.org/10.1038/sdata.2017.40

Lewis, L.D., Setsompop, K., Rosen, B.R., Polimeni, J.R., 2018. Stimulus-dependent hemodynamic response timing across the human subcortical-cortical visual pathway identified through high spatiotemporal resolution 7T fMRI. NeuroImage 181, 279–291. https://doi.org/10.1016/j.neuroimage.2018.06.056

Liao, C.H., Worsley, K.J., Poline, J.-B., Aston, J.A.D., Duncan, G.H., Evans, A.C., 2002. Estimating the Delay of the fMRI Response. NeuroImage 16, 593–606. https://doi.org/10.1006/nimg.2002.1096

Lindquist, M.A., Meng Loh, J., Atlas, L.Y., Wager, T.D., 2009. Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. NeuroImage, Mathematics in Brain Imaging 45, S187–S198. https://doi.org/10.1016/j.neuroimage.2008.10.065

Lloyd, B., de Voogd, L.D., Mäki-Marttunen, V., Nieuwenhuis, S., 2023. Pupil size reflects activation of subcortical ascending arousal system nuclei during rest. eLife 12, e84822. https://doi.org/10.7554/eLife.84822

Mathot, S., 2018. Pupillometry: Psychology, Physiology, and Function. Journal of Cognition 1. https://doi.org/10.5334/joc.18

McDougal, D.H., Gamlin, P.D., 2015. Autonomic Control of the Eye. Comprehensive Physiology 5, 439–473. https://doi.org/10.1002/j.2040-4603.2015.tb00603.x

McMahon, E., Bonner, M.F., Isik, L., 2023. Hierarchical organization of social action features along the lateral visual pathway. Current Biology 33, 5035-5047.e8. https://doi.org/10.1016/j.cub.2023.10.015

Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain structure & function 214, 655–67. https://doi.org/10.1007/s00429-010-0262-0

Murphy, P.R., O'Connell, R.G., O'Sullivan, M., Robertson, I.H., Balsters, J.H., 2014. Pupil diameter covaries with BOLD activity in human locus coeruleus. Human Brain Mapping 35, 4140–4154. https://doi.org/10.1002/hbm.22466

Nastase, S.A., Gazzola, V., Hasson, U., Keysers, C., 2019. Measuring shared responses across subjects using intersubject correlation. Soc Cogn Affect Neurosci 14, 667–685. https://doi.org/10.1093/scan/nsz037

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J.L., 2011. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Current Biology 21, 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I.P., Hari, R., Sams, M., 2012. Emotions promote social interaction by synchronizing brain activity across individuals. Proceedings of the National Academy of Sciences 109, 9599–9604. https://doi.org/10.1073/pnas.1206095109

Pais-Roldán, P., Takahashi, K., Sobczak, F., Chen, Y., Zhao, X., Zeng, H., Jiang, Y., Yu, X., 2020. Indexing brain state-dependent pupil dynamics with simultaneous fMRI and optical fiber calcium recording. Proc Natl Acad Sci U S A 117, 6875–6882. https://doi.org/10.1073/pnas.1909937117

Pedregosa, F., Eickenberg, M., Ciuciu, P., Thirion, B., Gramfort, A., 2015. Data-driven HRF estimation for encoding and decoding models. NeuroImage 104, 209–220. https://doi.org/10.1016/j.neuroimage.2014.09.060

Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., Jacob, Y., Atzil, S., Admon, R., Bleich-Cohen, M., Maron-Katz, A., Hendler, T., Barrett, L.F., 2016. Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. Cogn Affect Behav Neurosci 16, 709–723. https://doi.org/10.3758/s13415-016-0425-4

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., Saxe, R., 2018. Development of the social brain from age three to twelve years. Nat Commun 9, 1–12. https://doi.org/10.1038/s41467-018-03399-2

Saarimäki, H., 2021. Naturalistic Stimuli in Affective Neuroimaging: A Review. Front. Hum. Neurosci. 15. https://doi.org/10.3389/fnhum.2021.675068

Sadaghiani, S., D'Esposito, M., 2015. Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. Cereb Cortex 25, 2763–2773. https://doi.org/10.1093/cercor/bhu072

Samara, A., Zada, Z., Vanderwal, T., Hasson, U., Nastase, S.A., 2025. Cortical language areas are coupled via a soft hierarchy of model-based linguistic features. https://doi.org/10.1101/2025.06.02.657491

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T., 2018. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cerebral cortex (New York, N.Y. : 1991) 28, 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schneider, M., Hathway, P., Leuchs, L., Sämann, P.G., Czisch, M., Spoormaker, V.I., 2016. Spontaneous pupil dilations during the resting state are associated with activation of the salience network. NeuroImage 139, 189–201. https://doi.org/10.1016/j.neuroimage.2016.06.011

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neuroscience & Biobehavioral Reviews 42, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Schurz, M., Radua, J., Tholen, M.G., Maliske, L., Margulies, D.S., Mars, R.B., Sallet, J., Kanske, P., 2021. Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. Psychological Bulletin 147, 293–327. https://doi.org/10.1037/bul0000303

Shine, J.M., Hearne, L.J., Breakspear, M., Hwang, K., Müller, E.J., Sporns, O., Poldrack, R.A., Mattingley, J.B., Cocchi, L., 2019. The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei. Neuron 104, 849-855.e3. https://doi.org/10.1016/j.neuron.2019.09.002

Siegel, J.S., Snyder, A.Z., Ramsey, L., Shulman, G.L., Corbetta, M., 2016. The effects of hemodynamic lag on functional connectivity and behavior after stroke. J Cereb Blood Flow Metab 36, 2162–2176. https://doi.org/10.1177/0271678X15614846

Sohn, W., Di, X., Liang, Z., Zhang, Z., Biswal, B.B., 2024. Explorations of using a convolutional neural network to understand brain activations during movie watching. https://doi.org/10.1101/2024.01.20.576341

Sonkusare, S., Breakspear, M., Guo, C., 2019. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. Trends in Cognitive Sciences 23, 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Spunt, R.P., Adolphs, R., 2014. Validating the Why/How contrast for functional MRI studies of Theory of Mind. Neuroimage 99, 301–311. https://doi.org/10.1016/j.neuroimage.2014.05.023

Taylor, A.J., Kim, J.H., Ress, D., 2018. Characterization of the hemodynamic response function across the majority of human cerebral cortex. NeuroImage 173, 322–331. https://doi.org/10.1016/j.neuroimage.2018.02.061

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273–89. https://doi.org/10.1006/nimg.2001.0978

Vanderwal, T., Eilbott, J., Castellanos, F.X., 2019. Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. Developmental Cognitive Neuroscience 36, 100600. https://doi.org/10.1016/j.dcn.2018.10.004

West, K.L., Zuppichini, M.D., Turner, M.P., Sivakolundu, D.K., Zhao, Y., Abdelkarim, D., Spence, J.S., Rypma, B., 2019. BOLD hemodynamic response function changes significantly with healthy aging. NeuroImage 188, 198–207. https://doi.org/10.1016/j.neuroimage.2018.12.012

Wheatley, T., Milleville, S.C., Martin, A., 2007. Understanding Animate Agents: Distinct Roles for the Social Network and Mirror System. Psychol Sci 18, 469–474. https://doi.org/10.1111/j.1467-9280.2007.01923.x

Yellin, D., Berkovich-Ohana, A., Malach, R., 2015. Coupling between pupil fluctuations and resting-state fMRI uncovers a slow build-up of antagonistic responses in the human cortex. NeuroImage 106, 414–427. https://doi.org/10.1016/j.neuroimage.2014.11.034

Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E.D., Grafman, J., 2009. The Neural Basis of Human Social Values: Evidence from Functional MRI. Cereb Cortex 19, 276–283. https://doi.org/10.1093/cercor/bhn080