Between-movie variability severely limits generalizability of "naturalistic" neuroimaging

Simon Leipold [1,2], Rajat Ravi Rao [1,3], Jan-Mathijs Schoffelen [1], Sara Bögels [1,4], & Ivan Toni [1]

[1] Donders Institute for Brain, Cognition, and Behaviour, Radboud University

[2] Department of Humanities, Social and Political Sciences, ETH Zürich

[3] Cognitive Psychology, Leiden University

[4] Centre for Language Studies, Radboud University

Corresponding authors:

Simon Leipold & Ivan Toni
Radboud University
Donders Institute for Brain, Cognition, and Behaviour
Centre for Cognitive Neuroimaging
Kapittelweg 29
6525 EN Nijmegen
The Netherlands

Emails: simon.leipold@gess.ethz.ch; ivan.toni@donders.ru.nl

Keywords: inter-subject correlation (ISC), neural synchronization, movie-watching fMRI, individual differences

Abstract

“Naturalistic imaging” paradigms, where participants watch movies during fMRI, have gained popularity over the past two decades. Many movie-watching studies measure inter-subject correlation (ISC), which refers to the correlation between participants' neural activation time series. Previous research has focused on explaining ISC differences during movie-watching based on individual states and traits, such as social distance, personality, and political orientation. For example, friends show higher ISC than strangers while watching movies.

However, movies are not natural categories but cultural artifacts that evoke varying levels of ISC depending on content, directing style, or editing methods. This raises questions about how much trait- or state-like differences in ISC depend on the specific movies chosen, potentially limiting the generalizability of findings across different movies.

Here, we used an fMRI dataset of 112 participants watching eight animated movies to (a) quantify between-movie variability in ISC across the brain and (b) assess the implications for the generalizability of trait- or state-like effects on ISC. We found substantial between-movie variability in ISC, with this variability differing across brain regions. Crucially, brain regions with the highest ISC exhibited the greatest variability, indicating that trait- or state-like differences in ISC from one movie may not generalize to others.

We conclude that variability between movies limits the generalizability of trait- or state-like ISC differences. Using a specific movie in neuroscience should be treated similarly to using a particular task, requiring a comparable characterization of the constituent cognitive elements. Broad generalizations about “naturalistic imaging” or “movie watching” are not warranted.

Introduction

In the last two decades, “naturalistic” paradigms have gained enormous popularity in neuroscience [1–6]. These paradigms involve subjects viewing movies [e.g., 7,8] or listening to narratives [e.g., 9–11] or music [12,13] while undergoing functional brain imaging, most commonly functional magnetic resonance imaging (fMRI). It has been argued that such naturalistic paradigms more closely resemble real-life experiences [14], thus overcoming the lack of “ecological validity” of cognitively controlled neuroscience experiments. Namely, the simple and arbitrarily ordered stimuli used in traditional neuroscience experiments might limit the generalizability of their findings to real-life settings [15]. However, to date, the generalizability of the findings generated by “naturalistic” paradigms has been largely overlooked, despite their reliance on culturally contingent and idiosyncratic artifacts like movies and literary narratives.

The most popular method for analyzing fMRI data from naturalistic paradigms is inter-subject correlation (ISC) analysis [7]. In ISC analysis, the fMRI time series of a brain region (e.g., a single voxel or brain atlas parcel) in one subject is used as a model for the fMRI time series in the same region of another subject [16]: Calculating the correlation coefficient of the two time series yields a measure of synchronization of brain activation between the different subjects [17]. Summary statistics, such as the average correlation between all possible pairs of subjects, are calculated to express the level of synchronization across all participants per brain region. These summary statistics themselves are commonly referred to as ISC.

Prior work has demonstrated that many naturalistic materials, particularly movies, evoke strong ISC across the cortex, with particularly strong synchronization in visual and auditory brain regions [7,17]. The spatial distribution of ISC across the brain is highly reliable when subjects view the same movie more than once [2]. An advantage of ISC over general linear model analysis is that it is completely data-driven with no need to build regressors of specific event types, such as events marking the appearance of faces during the movie. Crucially, ISC is *stimulus-driven*, as activation changes that are not consistently related to processing of the naturalistic material across participants remain largely uncorrelated and do not contribute to the ISC [16,17]. However, it remains unclear whether different naturalistic materials, for example different movie clips, evoke consistent patterns of ISC.

In recent years, research employing naturalistic paradigms has increasingly focused on explaining state-like and trait-like differences in neural synchronization between subjects. In a

typical study, a measure of neural synchronization, most commonly ISC, is assessed while subjects watch a movie or listen to a narrative. Neural synchronization is compared between (sub-)groups of subjects classified based on certain characteristics, such as psychological states, traits, or clinical diagnoses. For example, prior work suggests that subjects with similar psychological perspectives, such as watching a movie from the perspective of a “detective” versus an “interior decorator” [18], or having similar beliefs about the characters in a story [19], show higher ISC than subjects with different perspectives or beliefs. Groups of subjects that reach consensus about the interpretation of ambiguous movie clips show higher ISC than groups that do not reach consensus [20]. Subjects who are friends show higher ISC than subjects who are further away from each other in a social network [21]. Subjects who are sisters show higher ISC than friends [22], and married couples show higher ISC than random couples [23]. In the clinical domain, alterations in ISC during movie watching have been reported in individuals with autism [24,25], psychosis [26], depression [27], and social anxiety [28], among others. This research line was recently extended within the framework of inter-subject representational similarity analysis (IS-RSA), which allows for a dimensional assessment of the relation between subjects’ neural synchronization patterns and subjects’ behavioral characteristics [29,30]. For example, subjects who are similar in age [31–33], personality traits [34,35], political orientation [36–39], sexual preferences [40], or reading ability [41] show higher neural synchronization than subjects who are dissimilar in these characteristics while watching a movie or listening to a narrative.

The majority of studies investigating state- or trait-like differences in ISC have drawn conclusions from the presentation of a single movie or narrative to the subjects. However, a few studies presenting more than one movie suggest that different movies might evoke different patterns of neural synchronization across the brain [20,23,28,39]. In other words, the same group of subjects might have higher synchronization than another group in brain region A for movie #1, higher neural synchronization in region B during movie #2, and no significant synchronization differences for movie #3.

By their very nature, movies are spatially and temporally complex, as well as highly multidimensional [7,42]. Different movies vary along multiple features. Considering that the main measure of neural synchronization, ISC, is strongly stimulus-driven [16,17], different movies might lead to different state- or trait-like effects in ISC. Indeed, the extraordinary sensitivity of ISC to varying features of movies was already highlighted in the earliest studies using naturalistic paradigms [2,7]. Notably, it has been suggested that ISC could be a useful tool for directors or

producers of movies to assess how engaging their movies are to an audience [43], as feature films evoke much stronger ISC than filmed segments of reality [43] or educational movies [33].

Surprisingly, this systematic variation in ISC between movies has been neglected when considering state- and trait-related ISC effects. More precisely, it is unclear how spatial patterns of ISC vary from movie to movie, a crucial element when considering the scope of findings based on a single movie [5,6]. How generalizable or “ecologically valid” are naturalistic paradigms in neuroscience, if state- or trait-like differences in ISC are bound to the movie being shown?

Here, we leverage a dataset of 112 adults watching eight movies each [44] to (i) quantify between-movie variability in ISC across the brain and (ii) assess the potential consequences of this variability in ISC for state- or trait-like differences in ISC (**Figure 1**). Specifically, we assessed how between-movie variability affects association between a state-like characteristic, similarity in semantic judgments of novel objects, and a neural synchronization characteristic, measured using ISC.

We find substantial between-movie variability in ISC across the whole brain, with the extent of variability differing strongly across brain regions: regions with higher levels of ISC showed higher variability. Crucially, applying IS-RSA for each movie separately revealed strongly varying associations of subjects' similarity in semantic judgments with neural synchronization. The localization of effects was different for each movie, which puts in doubt the generalizability of findings obtained from a single or a few movies.

Results

Between-movie variability in ISC is substantial

Our first aim was to quantify the variability in ISC evoked by different movies across the brain. We analyzed fMRI data from 112 adults (76 females; mean age = 22.72) from a published dataset [44]. The data was acquired while subjects watched eight animated movies (detailed in **Supplementary Table 1**). Animated movies are an ideal limiting case for assessing movie-related variability in ISC, as they are stylistically and thematically similar. For each subject and movie, we extracted a preprocessed and spatially averaged fMRI time series from each of the 210 cortical parcels of the Brainnetome atlas [45]. For each parcel, we calculated the ISC between all subject pairs by computing the Pearson correlation coefficient between the extracted fMRI time series. Subsequently, we averaged the pairwise correlation coefficients to obtain a mean ISC value per parcel.

As shown in **Figure 2**, the overall spatial distribution of ISC across the brain evoked by each of the eight movies resembled the spatial pattern found in numerous studies using movies [2,6,7]. The strongest ISC was evoked in the extended visual system, including face-sensitive areas in the fusiform gyrus, and the auditory system, including Heschl's gyrus and the planum temporale on the superior temporal plane. Additionally, there was strong ISC in higher-order areas in the temporal, parietal, and frontal cortices. The ISC values for each parcel and movie are available online (https://dx.doi.org/10.17605/OSF.IO/S78WU). At first glance, the overall pattern of ISC looked qualitatively similar across movies. In prior work, these qualitative similarities have been taken as evidence that different naturalistic materials lead to similar neural synchronization across the brain [46]. However, upon closer visual inspection of the ISC patterns in **Figure 2**, it becomes apparent that different movies did not elicit comparable ISC. For example, Movie #3 seemed to elicit generally higher ISC values than Movie #5. Crucially, the spatial *patterns* of ISC values across movies are clearly different.

Following up on these qualitative observations, we next quantified the extent of variability in ISC between movies. First, we investigated if the extent of whole-brain ISC varies between movies. To do this, we averaged the ISC values across all cortical parcels to obtain one whole-brain ISC value per subject pair and movie. This pair-wise approach is one of the standard procedures when computing ISC [16]. Then, we submitted the whole-brain ISC values to a repeated-measures analysis of variance (ANOVA) with Movie as the within-subject factor. To meet the

assumptions of the ANOVA, we used the values from a subset of subject pairs, ensuring each subject uniquely contributed to a single subject pair. The results of the ANOVA revealed a significant main effect of Movie ($F(7,385) = 4.65$, $p < 0.001$, generalized eta-squared [η^2_G] = 0.048), indicating substantial variability in whole-brain ISC across movies. **Figure 3A** visualizes the means and standard errors of the whole-brain ISC values for each movie separately. Notably, the visualization shows that the differences in ISC values between movies are not driven by one particular movie being different from all the others; rather, ISC values consistently differ across all movies.

Next, we investigated the spatial distribution of ISC variability by quantifying the between-movie variability in ISC values across all parcels of the cortex. We conducted a repeated-measures ANOVA with Movie as the within-subject factor for each of the 210 cortical parcels defined by the Brainnetome atlas. We found a statistically significant main effect of Movie (family-wise error (FWE)-adjusted $p < 0.05$) for 96 of the 210 cortical parcels (45.7%). The strongest variability was found in the primary and associative visual cortex, as well as the primary and associative auditory cortex. However, there was also substantial variability in higher-order areas, such as the precuneus and inferior parietal lobule. The statistical values quantifying variability for each parcel are available online (https://dx.doi.org/10.17605/OSF.IO/S78WU). **Figure 3B** shows the spatial distribution of between-movie ISC variability across the brain.

Brain regions with stronger ISC show higher between-movie variability in ISC

We followed up on the finding of widespread and substantial between-movie variability in ISC across the cortex by inspecting brain regions that exhibited particularly large and small variability. We discovered that brain regions showing high variability, such as the superior temporal cortex, also had relatively higher levels of ISC. Conversely, brain regions with low variability, such as the parahippocampal gyrus, had lower levels of ISC (**Figure 3C**). To quantify this relationship, we calculated the Pearson correlation coefficient between the level of ISC and between-movie variability of ISC across the 210 brain regions. As shown in **Figure 3D**, ISC levels and between-movie variability were strongly positively correlated ($r = 0.70$, $t(208) = 14.21$, $p < 0.001$, $R^2 = 0.49$), indicating that brain regions with stronger ISC showed higher between-movie variability in ISC. Thus, the brain regions that synchronize most across subjects when viewing movies are the most strongly affected by variability in ISC between movies.

Between-movie variability in ISC affects generalizability of state- and trait-like differences in ISC

Our next objective was to assess consequences of between-movie variability in ISC for the generalizability of state- or trait-like differences in ISC. Our dataset encompasses behavioral data on subjects' conceptual-semantic representations of novel objects ("Fribbles", **Figure 1C**), which were estimated in a *Features task* [44,47]. In this task, subjects used a linear visual analog scale to report how well each novel object matched 29 different features (e.g., pointy, symmetrical, human, related to movement; full list in **Supplementary Table 2**; see **Supplementary Methods** for details on the Features task). Prior work has shown that subjects with similar interpretations of a narrative show higher neural synchronization than subjects with different interpretations [19]. Furthermore, subjects who share their interpretations of an animation of abstract geometric shapes show high neural synchronization during narrative listening in the default-mode network [48]. Inspired by these previous studies, we hypothesized that subjects with similar conceptual representations of novel objects would show higher ISC during movie watching. As these particular movies were selected to contain categories of objects that were mentioned often in pilot tests to be associated with the novel objects [44], we anticipated a "spill-over" effect from representations of novel objects to naturalistic viewing of conventional objects.

We used IS-RSA to discover brain regions in which similarity in conceptual representations of novel objects (as measured by the Features task) across pairs of participants are associated with similarity in ISC during movie watching. In other words, we looked for brain regions that synchronize more between subjects that have similar semantic representations of the novel objects. For each movie separately, we performed regression-based IS-RSA using linear mixed-effects models [49] to estimate the association between subject similarity in semantic representations and ISC values during movie watching, while controlling for subjects' similarities in the names they gave the novel objects (*Naming task*) and similarities in age and sex. Linear mixed-effects models are commonly used for quantifying state- or trait-like differences in ISC [32,34,36,39,41,50,51]. If the effects of between-movie variability in ISC are negligible, we would expect a similar pattern of IS-RSA results across movies. By contrast, we would expect a distinct pattern of IS-RSA results for each movie if between-movie variability in ISC is consequential for assessing state- or trait-like differences in ISC.

As shown in **Figure 4**, the results of the IS-RSA revealed associations between semantic representations of novel objects and ISC in distinct brain regions for each movie. For Movie #1,

198 we found that similarities between subjects in semantic representations were statistically
199 significantly associated with neural synchronization between subjects in parcels within bilateral
200 insula (right hemisphere: p_{FWE} = 0.004 and left hemisphere: p_{FWE} = 0.007), a parcel in the left
201 inferior parietal lobule (p_{FWE} = 0.006), the left precentral gyrus (p_{FWE} = 0.007), the right superior
202 frontal gyrus (p_{FWE} = 0.02), the left inferior temporal gyrus (p_{FWE} = 0.03), and the right anterior
203 cingulate (p_{FWE} = 0.03). In contrast, for Movie #2, we found no statistically significant association
204 between semantic representations and neural synchronization between subjects. For Movie #3,
205 we found significant associations of similarities in semantic representations with ISC in the left
206 precuneus (p_{FWE} < 0.05), which did not overlap with any of the regions that were found for Movie
207 #1. For Movie #4, we found significant associations in bilateral middle frontal gyrus (right
208 hemisphere: p_{FWE} < 0.001 and left hemisphere: p_{FWE} = 0.03) and the right orbital gyrus (p_{FWE} =
209 0.02), which again did not overlap with regions found in Movie #1 or #3. Likewise, for Movie #5
210 through Movie #8, we found non-overlapping brain regions that showed an association between
211 semantic representations and neural synchronization except for a single parcel in the right
212 middle temporal gyrus, which showed an association both for Movie #5 (p_{FWE} = 0.004) and
213 Movie #6 (p_{FWE} < 0.001). Details on the statistically significant IS-RSA associations for all
214 movies are listed in **Supplementary Table 3**. Details on all parcels and movies are available
215 online (https://dx.doi.org/10.17605/OSF.IO/S78WU). Thus, between-movie variability in ISC has
216 detrimental downstream effects for generalizability of findings with one movie as these effects
217 likely do not generalize to other movies.

Discussion

Here, we quantified the extent of variability in ISC between different movies, and we assessed the consequences of this variability for investigations of state-like or trait-like differences in neural synchronization during movie watching. Our analyses revealed substantial between-movie variability in whole-brain ISC and statistically significant variability in almost half of all cortical regions. Between-movie variability in ISC affected not only early auditory and visual brain regions but also higher-order brain regions, including those belonging to the default-mode network. Using IS-RSA, we showed that associations between subjects' similarity in semantic-conceptual representations of novel objects and ISC during movie watching have a distinct spatial distribution in the brain for each of the movies, with minimal overlap between movies. Thus, findings of state-like and trait-like differences in ISC based on a single movie are not generalizable to different movies.

What features of a movie drive neural synchronization?

A major issue in using naturalistic material to investigate state-like and trait-like differences in neural synchronization is that it is unknown which features of a movie might drive the neural synchronization. As resting-state fMRI time series between different subjects are generally uncorrelated [7], there must be features of the naturalistic material that drive the neural synchronization. However, ever since the first studies used free viewing of movies as a paradigm in neuroscience [7,42], it has been clear that these stimuli are too complex and multidimensional to isolate specific features that drive neural synchronization [2,7,52]. Thus, we argue that even when movies are grouped by an experimenter-selected feature, for example "child-oriented animations", different movie clips within this specific "category" will still vary in a myriad of features that might differentially drive brain responses and lead to different neural synchronization between subjects.

Including the stimulus as a random effect in the statistical model

A potential solution to address the non-generalizability of findings derived from a single or a small number of movies is to include the movie as a "random effect" in the statistical model. Following Yarkoni [53], "random effects are used to model [...] variables that are assumed to be stochastically sampled from some underlying population and can vary across replications without meaningfully altering the research question." [53 p. 3]. With rare exceptions [39], studies on

248 state- or trait-like differences in ISC that used more than one naturalistic stimulus have typically
249 modeled the naturalistic materials as a fixed effect [e.g., 20] or ran separate models for each
250 stimulus, effectively treating the stimulus as fixed. Treating the stimulus as fixed limits the ability
251 to generalize findings beyond the specific stimulus set used in the study. This issue has long
252 been recognized as a problem in both neuroscience [54] and other scientific fields [53,55,56]. Thus, it
253 is possible that the phenomenon we describe here is a specific case of "treating the stimulus as
254 a fixed effect fallacy", and can, in theory, be resolved by including the stimulus as a random
255 effect in the statistical model. However, this approach raises important questions: What are we
256 generalizing across? Is it reasonable to assume that there are findings that generalize across a
257 category of "movie watching" or "narrative listening", even though different movie clips and
258 narratives are unique in many ways? Are different movies truly addressing the "same research
259 question", as they would when included as a random effect, or do they, in fact, each pose their
260 own unique research questions?

261 **Movies are akin to tasks and not rest**

262 Naturalistic paradigms have been proposed as an alternative to resting-state paradigms [5], and
263 state-like or trait-like differences in resting-state characteristics seem to be generalizable [57].
264 Movie watching overcomes many limitations of resting-state acquisitions. For example, findings
265 from movie-watching show higher test-retest reliability when subjects watch the same movie
266 multiple times [58]. Additionally, movies are well-suited for young children and pediatric
267 populations [6], as they are associated with higher attentional engagement and, importantly, less
268 head motion than resting-state conditions [31,59]. However, when different movies evoke unique
269 patterns of neural synchronization between subjects, as our results suggest, different movies
270 may be more akin to different tasks than to resting-state acquisitions. Consequently, studies
271 need to describe movies as carefully as it is usually done for tasks, despite the difficulties of
272 identifying which exact features are important [6]. A potential solution for evoking a state in
273 subjects that is closer to rest involves the use of specifically designed audiovisual materials,
274 such as *Inscapes*, which elicit more rest-like brain activation patterns than feature movies [59].

275 **Movies are not a naturalistic depiction of reality**

276 It has been argued that the rich and complex stimulation provided by a movie is much closer to
277 real-life perception compared to custom-designed stimuli used in laboratory settings [7]. In a
278 seminal study where subjects watched a *James Bond* movie, the paradigm was described as

subjects observing a "dynamic natural scene" [42]. However, a Hollywood-produced action movie is far from natural: at best, movies are an abstracted representation of daily-life experiences. They contain elements that are deliberately manipulated by filmmakers to capture and maintain the audience's attention (e.g., camera movements and zooms, editing techniques such as cuts, soundtracks). Indeed, researchers have noted early on that critically acclaimed movies evoke much higher ISC than simply watching actual natural scenes [43] or educational films [33]. While naturalistic paradigms may appear more ecologically valid than controlled experiments with abstract stimuli, it remains unclear to what extent ISC during movie-watching is largely driven by stylistic elements of movies, such as cutting and zooming.

Limits of generalizability of current findings and future directions

It remains an open question to what extent our findings are generalizable to all studies using naturalistic materials. For example, studies using control conditions such as scrambled movies [8,9] or scrambled music [12] may be less affected by the stimulus-specificity in ISC reported here. The logic behind the scrambled-conditions approach is that many (though likely not all) features of naturalistic materials that drive variable ISC across movie clips or music pieces are controlled between the real and scrambled conditions, except for the ones of interest to the researcher. However, if this assumption is not fully met – because it is difficult to match all features that variably drive ISC between the real and control conditions – stimulus-specificity could still pose a problem. Additionally, our findings may be less relevant for studies investigating specific types of events within naturalistic materials, such as those in memory or event segmentation studies using naturalistic materials [60–62]. In such cases, between-movie variability in ISC may have limited downstream consequences. The same might be true for studies employing voxel-wise modeling of naturalistic material [42,63–66].

It is likely that the adverse effects of between-stimulus variability on state- and trait-like differences in ISC are not unique to movies but may extend to other classes of naturalistic materials, such as narratives and music; further analysis methods, including inter-subject functional correlation [cf. 67]; as well as other neuroimaging modalities beyond fMRI, including magnetoencephalography (MEG), electroencephalography (EEG) [cf. 33], or functional near-infrared spectroscopy (fNIRS). However, these predictions require empirical confirmation in future work.

Conclusion

We find substantial between-movie variability in ISC across the cortex, leading to distinct spatial distributions of state-like differences in the brain for separate movies, with minimal overlap between movies. Consequently, findings of state-like and trait-like differences in ISC based on a single movie are not generalizable across different movies, raising questions about the validity of findings from previous studies that relied on a single or limited set of movies.

Methods

Subjects

We analyzed data from *N* = 112 subjects from the Communicative Alignment of Brain and Behaviour (CABB) dataset [44]. These subjects were the same as those included in the fMRI analysis in the dataset description paper by Eijk et al. [44] and did not include subjects who were excluded due to technical malfunctions, incomplete data, or poor data quality. For our main analysis, a repeated-measures ANOVA on ISC values of 56 subject pairs, we had 80% statistical power to detect differences in ISC between movies with an effect size of at least Cohen's $f > 0.51$. Statistical power was calculated using the *WebPower* package (version 0.9.2) [68] in *R* (version 4.1.0; RRID:SCR_001905).

fMRI movie task

Subjects viewed eight animated movies (see **Supplementary Table 1** for details), which were presented using Presentation software (version 20.2; RRID:SCR_002521). The movies were selected to include categories of objects frequently mentioned in pilot tests of the novel objects ("Fribbles", **Figure 1C**), such as humans, plants, tools, toys, and food. The movies were presented on a portion of the screen slightly below the center, on a black background with a height of 360 pixels and a width of 640 pixels, with a visual angle of approximately 9–11° vertically and 16–20° horizontally. The movies were played at 30 frames per second. Each movie was played in its entirety, except for the start and end (i.e., titles and credits), which were cut off to ensure no text was shown to subjects. The average duration of the movies was 4.1 minutes (range = 2.2-6.1 minutes), totaling 35 minutes, including 12-second breaks between subsequent movies. The movies were preceded by a filler video clip that lasted a few seconds. Subjects were instructed to simply attend to the movies and to lie still. The movies did not contain spoken language, but the original soundtrack was provided to subjects via earphones. Before scanning began, the sound levels were adjusted to each subject's individual preferences.

Imaging data acquisition

Magnetic resonance images were acquired using two 3T MAGNETOM magnetic resonance imaging scanners: Prisma and PrismaFit (Siemens AG, Healthcare Sector, Erlangen,

Germany). Blood oxygenation level dependent (BOLD) functional images were acquired using a multi-band 2D-echo planar imaging (EPI) sequence (Human Connectome Project) [69] with repetition time (TR) = 1000 ms, echo time (TE) = 34 ms, flip angle = 60º, 2mm isotropic resolution, field of view (FOV) = 208x208x132 mm. A multi-band acceleration factor of 6 was used in the slice direction, and no parallel imaging was applied in-plane. Phase encoding was applied in the anterior-posterior (AP) direction with a partial Fourier coverage of 7/8. About 2074 volumes were acquired during movie watching. Anatomical images were acquired with a T1-weighted 3D MPRAGE sequence, integrated parallel acquisition technique (iPAT) acceleration factor = 2, with TR = 2200 ms, TE = 2 ms, inversion time (TI) = 1100 ms, flip angle = 11°, 0.8 mm isotropic resolution, FOV = 256x240x166 mm; and with a T2-weighted turbo spin echo (TSE) sequence with variable flip angle, TR = 3200 ms, TE = 569 ms, echo spacing = 4.12 ms, turbo factor = 314, 0.8 mm isotropic resolution, and FOV = 256x240x166 mm. Raw imaging data were converted to Brain Imaging Data Structure (*BIDS*) format [70] using BIDScoin (version 1.5; RRID:SCR_022839) [71].

Imaging data preprocessing

Each subject's raw imaging data underwent a quality control assessment for excessive motion, susceptibility artifacts, or other sources of noise using *MRIQC* (version 23.1.0; RRID:SCR_022942). Imaging data preprocessing was implemented using *fMRIPrep* (version 20.2.7; RRID:SCR_016216) [72], which is based on *Nipype* (version 1.7.0; RRID:SCR_002502) [73]. The main preprocessing steps included segmentation, surface reconstruction, and spatial normalization of the T1-weighted anatomical image; and slice timing correction, susceptibility distortion correction, head-motion correction, and spatial normalization of the fMRI data. Details of the preprocessing pipeline are included in the **Supplementary Methods**.

fMRI data denoising

Each subject's preprocessed fMRI data underwent further denoising, implemented using *Nilearn* (version 0.9.2; RRID:SCR_001362) in *Python* (version 3.7.13; RRID:SCR_008394). The voxel-wise fMRI time series were detrended and low-pass filtered at 0.08 Hz. Next, the following confounds estimated by fMRIprep during preprocessing were regressed out from the voxel-wise time series: six motion parameters (three translations, three rotations) and their first temporal derivatives, as well as cerebrospinal fluid and white matter time series and their first temporal derivatives. Subsequently, the time series were standardized to have a mean of zero and unit

variance. After denoising, we segmented the fMRI time series and extracted a separate time series for each movie clip based on the onset and offset times recorded in the Presentation log files. The denoised fMRI data for each movie was then spatially smoothed with a 6 mm full-width-at-half-maximum three-dimensional Gaussian kernel.

Inter-subject correlation (ISC) analysis

To calculate parcel-wise ISC values, we first extracted a spatially averaged parcel-wise time series for each subject and movie from each of the 210 cortical parcels defined by the Brainnetome atlas [45] using *NLtools* (version 0.4.7) [74] in *Python*. For each movie and cortical parcel, we then calculated the ISC values for all possible subject pairs by computing the Pearson correlation coefficient between the parcel-wise fMRI time series of subject [i] and subject [j] of each pair [i,j] using *scikit-learn* (version 1.0.2; RRID:SCR_002577) in *Python*. To obtain a mean ISC value for each parcel, we averaged the ISC values across all subject pairs. Visualization of the ISC values in the brain was implemented using *Nilearn*.

Statistical analyses of ISC values

To quantify between-movie variability in ISC across the whole brain, we averaged the ISC values across all cortical parcels to obtain a single ISC value per subject pair and movie. Then, we entered the whole-brain ISC values into a repeated-measures ANOVA with Movie as the within-subject factor. We used data from 56 subject pairs, ensuring that each subject contributed only to a single ISC value per movie and that the assumptions of the repeated-measures ANOVA were not violated. Next, we repeated the ANOVA for each parcel separately to obtain a spatial distribution of between-movie variability across the brain. The ANOVAs were performed using the *ez* (version 4.4-0; RRID:SCR_020990) and *rstatix* (version 0.7.2; RRID:SCR_021240) packages in *R*.

To investigate whether brain regions with high levels of ISC also show high variability in ISC, we correlated the parcel-wise level of ISC with the parcel-wise between-movie variability in ISC. Specifically, we extracted the *F* values from the repeated-measures ANOVAs for each parcel as a measure of between-movie variability. We then calculated the Pearson correlation coefficient between the F values and the ISC values for each parcel in *R*.

Inter-subject representational similarity analysis (IS-RSA)

We conducted a regression-based IS-RSA [29,30] using linear mixed-effects models [39,49] to link subject similarity in conceptual representations of the novel objects, estimated in the *Features task*, with neural synchronization during movie watching, separately for each movie. The linear mixed-effects models were estimated using the *lme4* package (version 1.1-33; RRID:SCR_015654) in *R* (version 4.1.0; RRID:SCR_001905). We estimated a full model to simultaneously assess the main effects of subjects' similarities in conceptual representations on their ISC during movie watching, while controlling for subject similarity in lexical representations of the novel objects, estimated in a *Naming task*, and also controlling for subject similarities in sex and age. Details on the novel object stimuli, Features and Naming tasks, and the IS-RSA are given in the **Supplementary Methods**.

Code availability

Python and R code is available online (https://doi.org/10.5281/zenodo.13348506).

Acknowledgments

This work was supported by the Dutch Research Council (NWO; Gravitation Grant 024.001.006 to the Language in Interaction Consortium), by the Swiss National Science Foundation (SNSF; Grants 206557 and 222073 to Simon Leipold), and the European Research Council (ERC; Advanced Grant 101054559 to Ivan Toni).

References

1. Eickhoff, S. B., Milham, M. & Vanderwal, T. Towards clinical applications of movie fMRI. *NeuroImage* **217**, 116860 (2020).

2. Hasson, U., Malach, R. & Heeger, D. J. Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* **14**, 40–48 (2010).

3. Jääskeläinen, I. P., Sams, M., Glerean, E. & Ahveninen, J. Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage* **224**, 117445 (2021).

4. Redcay, E. & Moraczewski, D. Social cognition in context: A naturalistic imaging approach. *NeuroImage* **216**, 116392 (2020).

5. Sonkusare, S., Breakspear, M. & Guo, C. Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends Cogn. Sci.* **23**, 699–714 (2019).

6. Vanderwal, T., Eilbott, J. & Castellanos, F. X. Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Dev. Cogn. Neurosci.* **36**, 100600 (2019).

7. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject Synchronization of Cortical Activity during Natural Vision. *Science* **303**, 1634–1640 (2004).

8. Hasson, U., Yang, E., Vallines, I., Heeger, D. J. & Rubin, N. A Hierarchy of Temporal Receptive Windows in Human Cortex. *J. Neurosci.* **28**, 2539–2550 (2008).

9. Lerner, Y., Honey, C. J., Silbert, L. J. & Hasson, U. Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *J. Neurosci.* **31**, 2906–2915 (2011).

10. Willems, R. M., Nastase, S. A. & Milivojevic, B. Narratives for Neuroscience. *Trends Neurosci.* **43**, 271–273 (2020).

11. Wilson, S. M., Molnar-Szakacs, I. & Iacoboni, M. Beyond Superior Temporal Cortex: Intersubject Correlations in Narrative Speech Comprehension. *Cereb. Cortex* **18**, 230–242 (2008).

12. Abrams, D. A. *et al.* Inter-subject synchronization of brain responses during natural music listening. *Eur. J. Neurosci.* **37**, 1458–1469 (2013).

13. Alluri, V. *et al.* Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage* **59**, 3677–3689 (2012).

14. Nastase, S. A., Goldstein, A. & Hasson, U. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage* **222**, 117254 (2020).

15. Brunswik, E. Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.* **62**, 193–217 (1955).

16. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 667–685 (2019).

17. Nummenmaa, L., Lahnakoski, J. M. & Glerean, E. Sharing the social world via intersubject neural synchronisation. *Curr. Opin. Psychol.* **24**, 7–14 (2018).

18. Lahnakoski, J. M. *et al.* Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage* **100**, 316–324 (2014).

19. Yeshurun, Y. *et al.* Same Story, Different Story: The Neural Representation of Interpretive Frameworks. *Psychol. Sci.* **28**, 307–319 (2017).

20. Sievers, B., Welker, C., Hasson, U., Kleinbaum, A. M. & Wheatley, T. Consensus-building conversation leads to neural alignment. *Nat. Commun.* **15**, 3936 (2024).

466 21. Parkinson, C., Kleinbaum, A. M. & Wheatley, T. Similar neural responses predict friendship.
467 *Nat. Commun.* **9**, 332 (2018).

468 22. Bacha-Trams, M. *et al.* Sisterhood predicts similar neural processing of a film. *NeuroImage*
469 **297**, 120712 (2024).

470 23. Li, L. *et al.* Neural synchronization predicts marital satisfaction. *Proc. Natl. Acad. Sci.* **119**,
471 e2202515119 (2022).

472 24. Hasson, U. *et al.* Shared and idiosyncratic cortical activation patterns in autism revealed
473 under continuous real-life viewing conditions. *Autism Res.* **2**, 220–231 (2009).

474 25. Salmi, J. *et al.* The brains of high functioning autistic individuals do not synchronize with
475 those of others. *NeuroImage Clin.* **3**, 489–497 (2013).

476 26. Mäntylä, T. *et al.* Aberrant Cortical Integration in First-Episode Psychosis During Natural
477 Audiovisual Processing. *Biol. Psychiatry* **84**, 655–664 (2018).

478 27. Guo, C. C., Nguyen, V. T., Hyett, M. P., Parker, G. B. & Breakspear, M. J. Out-of-sync:
479 disrupted neural activity in emotional circuitry during film viewing in melancholic depression.
480 *Sci. Rep.* **5**, 11605 (2015).

481 28. Camacho, M. C. *et al.* Higher Intersubject Variability in Neural Response to Narrative Social
482 Stimuli Among Youth With Higher Social Anxiety. *J. Am. Acad. Child Adolesc. Psychiatry*
483 (2023) doi:10.1016/j.jaac.2023.08.020.

484 29. Finn, E. S. *et al.* Idiosynchrony: From shared responses to individual differences during
485 naturalistic neuroimaging. *NeuroImage* **215**, 116828 (2020).

486 30. Nummenmaa, L. *et al.* Emotions promote social interaction by synchronizing brain activity
487 across individuals. *Proc. Natl. Acad. Sci.* **109**, 9599–9604 (2012).

31. Cantlon, J. F. & Li, R. Neural Activity during Natural Viewing of Sesame Street Statistically Predicts Test Scores in Early Childhood. *PLOS Biol.* **11**, e1001462 (2013).

32. Moraczewski, D., Chen, G. & Redcay, E. Inter-subject synchrony as an index of functional specialization in early childhood. *Sci. Rep.* **8**, 2252 (2018).

33. Petroni, A. *et al.* The Variability of Neural Responses to Naturalistic Videos Change with Age and Sex. *eNeuro* **5**, (2018).

34. Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A. & Constable, R. T. Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nat. Commun.* **9**, 2043 (2018).

35. Matz, S. C., Hyon, R., Baek, E. C., Parkinson, C. & Cerf, M. Personality similarity predicts synchronous neural responses in fMRI and EEG data. *Sci. Rep.* **12**, 14325 (2022).

36. de Bruin, D., van Baar, J. M., Rodríguez, P. L. & FeldmanHall, O. Shared neural representations and temporal segmentation of political content predict ideological similarity. *Sci. Adv.* **9**, eabq5920 (2023).

37. Katabi, N. *et al.* Deeper Than You Think: Partisanship-Dependent Brain Responses in Early Sensory and Motor Brain Regions. *J. Neurosci.* **43**, 1027–1037 (2023).

38. Leong, Y. C., Chen, J., Willer, R. & Zaki, J. Conservative and liberal attitudes drive polarized neural responses to political content. *Proc. Natl. Acad. Sci.* **117**, 27731–27739 (2020).

39. van Baar, J. M., Halpern, D. J. & FeldmanHall, O. Intolerance of uncertainty modulates brain-to-brain synchrony during politically polarized perception. *Proc. Natl. Acad. Sci.* **118**, e2022491118 (2021).

40. Chen, P.-H. A., Jolly, E., Cheong, J. H. & Chang, L. J. Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *NeuroImage* **216**, 116851 (2020).

41. Jangraw, D. C. *et al.* Inter-subject correlation during long narratives reveals widespread neural correlates of reading ability. *NeuroImage* **282**, 120390 (2023).

42. Bartels, A. & Zeki, S. Functional brain mapping during free viewing of natural scenes. *Hum. Brain Mapp.* **21**, 75–85 (2004).

43. Hasson, U. *et al.* Neurocinematics: The Neuroscience of Film. *Projections* **2**, 1–26 (2008).

44. Eijk, L. *et al.* The CABB dataset: A multimodal corpus of communicative interactions for behavioural and neural analyses. *NeuroImage* **264**, 119734 (2022).

45. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).

46. Nastase, S. A. *et al.* The “Narratives” fMRI dataset for evaluating models of naturalistic language comprehension. *Sci. Data* **8**, 250 (2021).

47. Binder, J. R. *et al.* Toward a brain-based componential semantic representation. *Cogn. Neuropsychol.* **33**, 130–174 (2016).

48. Nguyen, M., Vanderwal, T. & Hasson, U. Shared understanding of narratives is correlated with shared neural responses. *NeuroImage* **184**, 161–170 (2019).

49. Chen, G., Taylor, P. A., Shin, Y.-W., Reynolds, R. C. & Cox, R. W. Untangling the relatedness among correlations, Part II: Inter-subject correlation group analysis through linear mixed-effects modeling. *NeuroImage* **147**, 825–840 (2017).

50. Baek, E. C. *et al.* In-degree centrality in a social network is linked to coordinated neural activity. *Nat. Commun.* **13**, 1118 (2022).

51. Baek, E. C. *et al.* Lonely Individuals Process the World in Idiosyncratic Ways. *Psychol. Sci.* **34**, 683–695 (2023).

52. Rust, N. C. & Movshon, J. A. In praise of artifice. *Nat. Neurosci.* **8**, 1647–1650 (2005).

53. Yarkoni, T. The generalizability crisis. *Behav. Brain Sci.* **45**, e1 (2022).

54. Westfall, J., Nichols, T. & Yarkoni, T. Fixing the stimulus-as-fixed-effect fallacy in task fMRI. *Wellcome Open Res.* **1**, (2017).

55. Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *J. Verbal Learn. Verbal Behav.* **12**, 335–359 (1973).

56. Coleman, E. B. Generalizing to a Language Population. *Psychol. Rep.* **14**, 219–226 (1964).

57. Gratton, C. *et al.* Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron* **98**, 439-452.e5 (2018).

58. Wang, J. *et al.* Test–retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Hum. Brain Mapp.* **38**, 2226–2241 (2017).

59. Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C. & Castellanos, F. X. Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage* **122**, 222–232 (2015).

60. Baldassano, C. *et al.* Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709-721.e5 (2017).

61. The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *J. Neurosci.* **38**, 10057–10068 (2018).

62. Chen, J. *et al.* Shared memories reveal shared structure in neural activity across individuals. *Nat. Neurosci.* **20**, 115–125 (2017).

63. Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).

64. Jacoby, N., Bruneau, E., Koster-Hale, J. & Saxe, R. Localizing Pain Matrix and Theory of Mind networks with both verbal and non-verbal stimuli. *NeuroImage* **126**, 39–48 (2016).

65. Richardson, H. Development of brain networks for social functions: Confirmatory analyses in a large open source dataset. *Dev. Cogn. Neurosci.* **37**, 100598 (2019).

66. Richardson, H., Lisandrelli, G., Riobueno-Naylor, A. & Saxe, R. Development of the social brain from age three to twelve years. *Nat. Commun.* **9**, 1027 (2018).

67. Ye, M., Liu, J., Guan, Y., Ma, H. & Tian, L. Are inter-subject functional correlations consistent across different movies? *Brain Imaging Behav.* **17**, 44–53 (2023).

68. Zhang, Z. & Yuan, K.-H. *Practical Statistical Power Analysis Using Webpower and R*. (ISDSA Press, 2018). doi:10.35566/power.

69. Uğurbil, K. *et al.* Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage* **80**, 80–104 (2013).

70. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016).

71. Zwiers, M. P., Moia, S. & Oostenveld, R. BIDScoin: A User-Friendly Application to Convert Source Data to Brain Imaging Data Structure. *Front. Neuroinformatics* **15**, (2022).

72. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).

73. Gorgolewski, K. J. *et al.* Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Front. Neuroinformatics* **5**, (2011).

74. Chang, L. *et al.* cosanlab/nltools: 0.4.7. Zenodo https://doi.org/10.5281/zenodo.7015135 (2022).

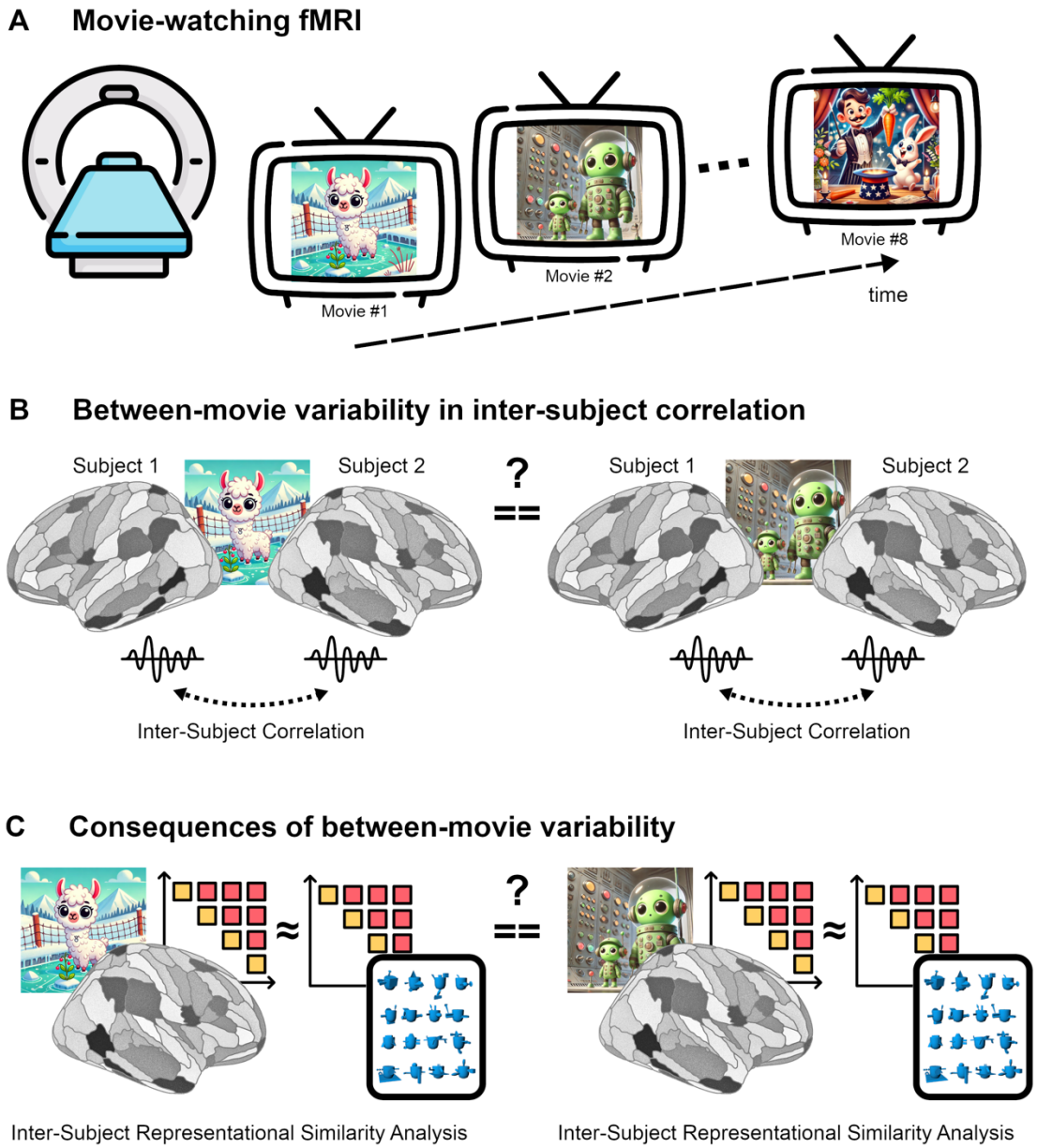


Figure 1. Overview.
A) Subjects (*N* = 112) viewed eight animated movies during fMRI scanning. The movies were selected to include categories of objects such as humans, plants, tools, toys, and food. **B)** We quantified the variability in inter-subject correlation (ISC) between movies across the brain. For each subject and movie, we extracted a preprocessed and spatially averaged fMRI time series from 210 cortical parcels. For each parcel, we calculated the ISC between all subject pairs by computing the Pearson correlation coefficient between the extracted fMRI time series. **C)** We assessed consequences of between-movie variability in ISC for the generalizability of findings of state- or trait-like differences in ISC. For this, we related similarities in subjects' conceptual-semantic representations of novel objects ("Fribbles", in blue), to neural synchronization between subjects during movie watching, for each movie separately.

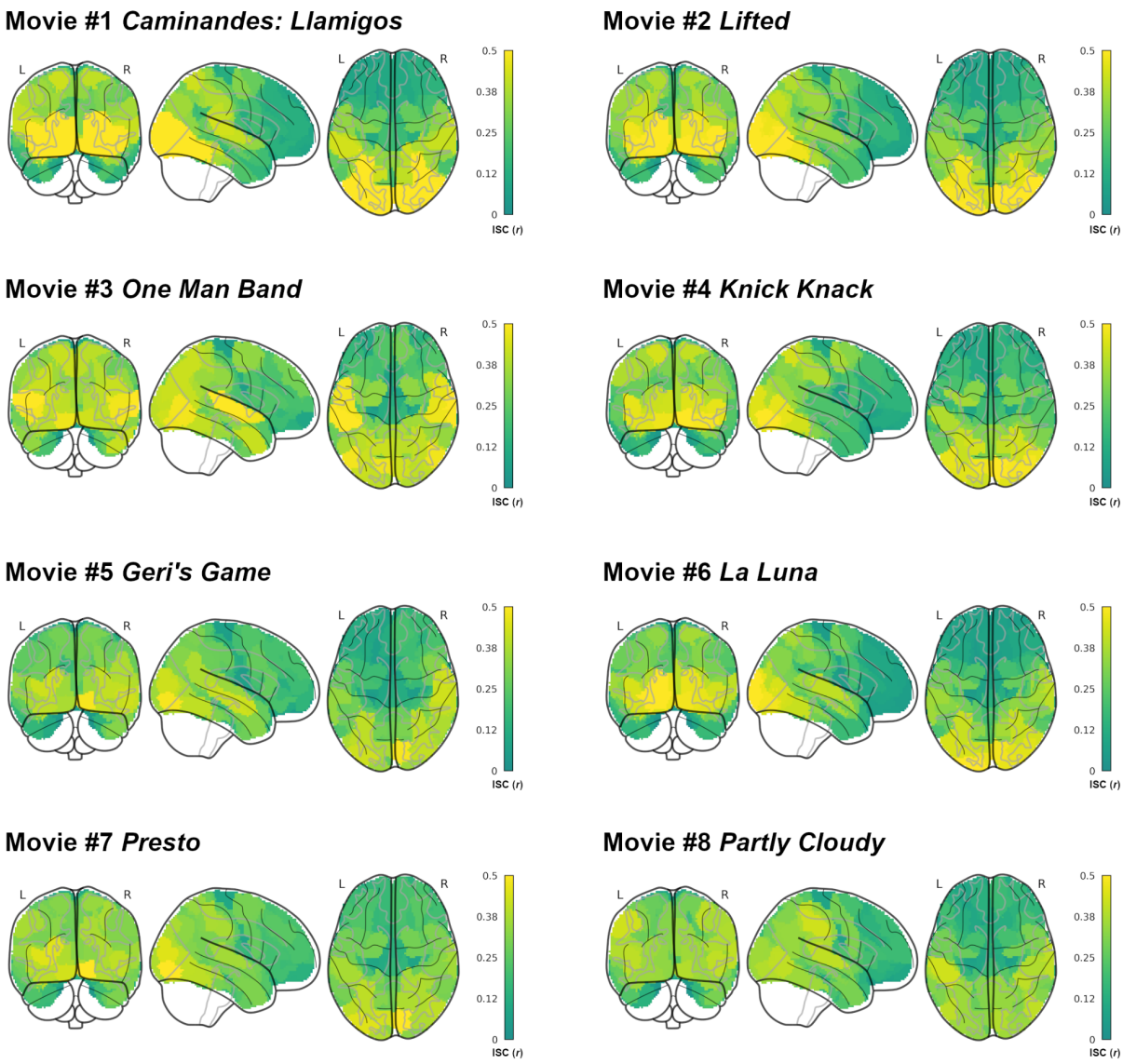


Figure 2. Inter-subject correlation for separate movies.
Neural synchronization between subjects looks qualitatively similar across movies but the specifics differ: Different movies did not elicit ISC to the same extent. For example, Movie #3 seemed to elicit generally higher ISC values than Movie #5. Importantly, the spatial patterns of ISC values across movies are not identical.

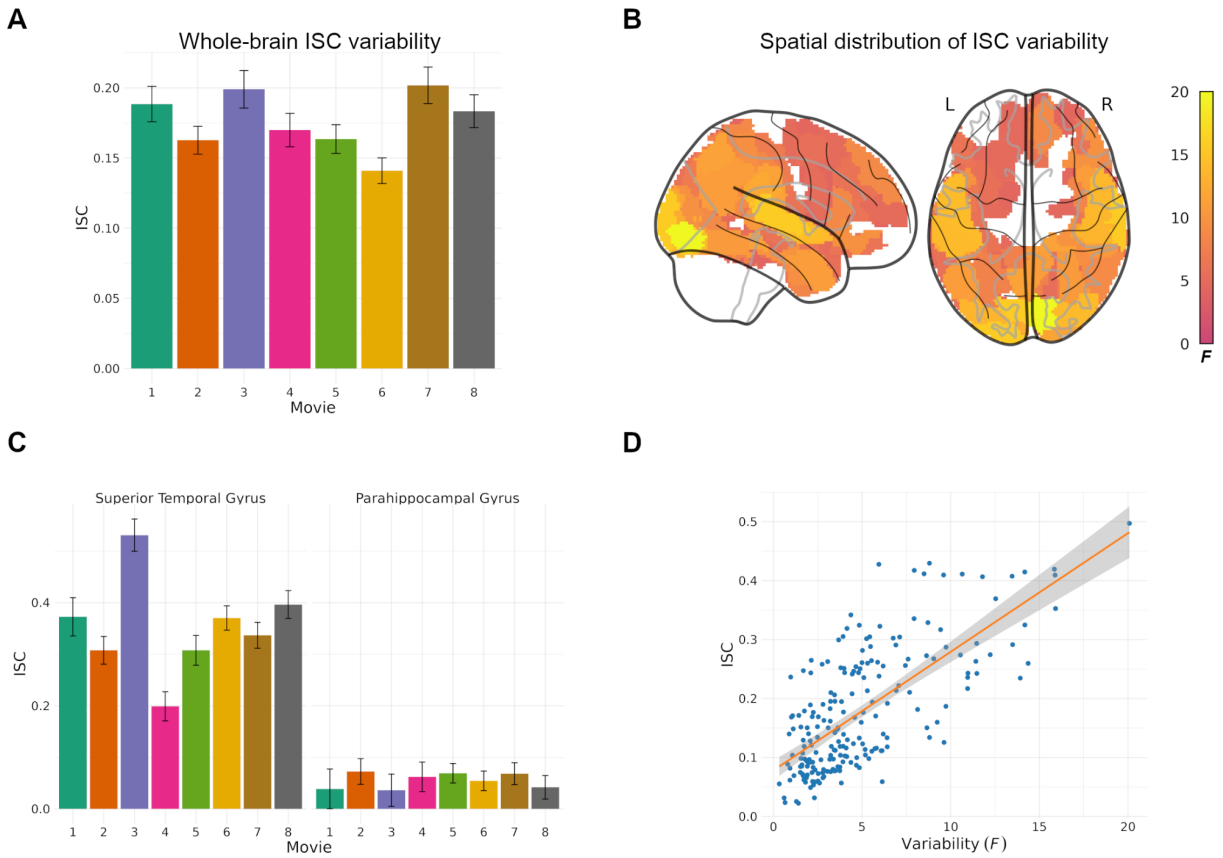


Figure 3. Variability in inter-subject correlation across movies.
A) Means (bar plot) and standard errors of the mean (error bar) of whole-brain ISC values for each of the eight movies. Whole-brain ISC values were obtained by averaging the ISC values across all cortical parcels for each subject and movie. The results of the ANOVA revealed a significant main effect of Movie, indicating substantial variability in whole-brain ISC across movies. **B)** Spatial distribution of between-movie ISC variability across the brain. We found a statistically significant main effect of Movie for 96 of the 210 cortical parcels (45.7%). The strongest variability was found in the primary and associative visual cortices, as well as the primary and associative auditory cortex. However, there was also substantial variability in higher-order areas, such as the precuneus and inferior parietal lobule. **C)** Brain regions showing high variability, such as the superior temporal gyrus, also had relatively higher levels of ISC. Conversely, brain regions with low variability, such as the parahippocampal gyrus, had lower levels of ISC. **D)** ISC levels (y-axis) and between-movie variability (x-axis) were strongly positively correlated, indicating that brain regions with stronger ISC showed higher between-movie variability in ISC.

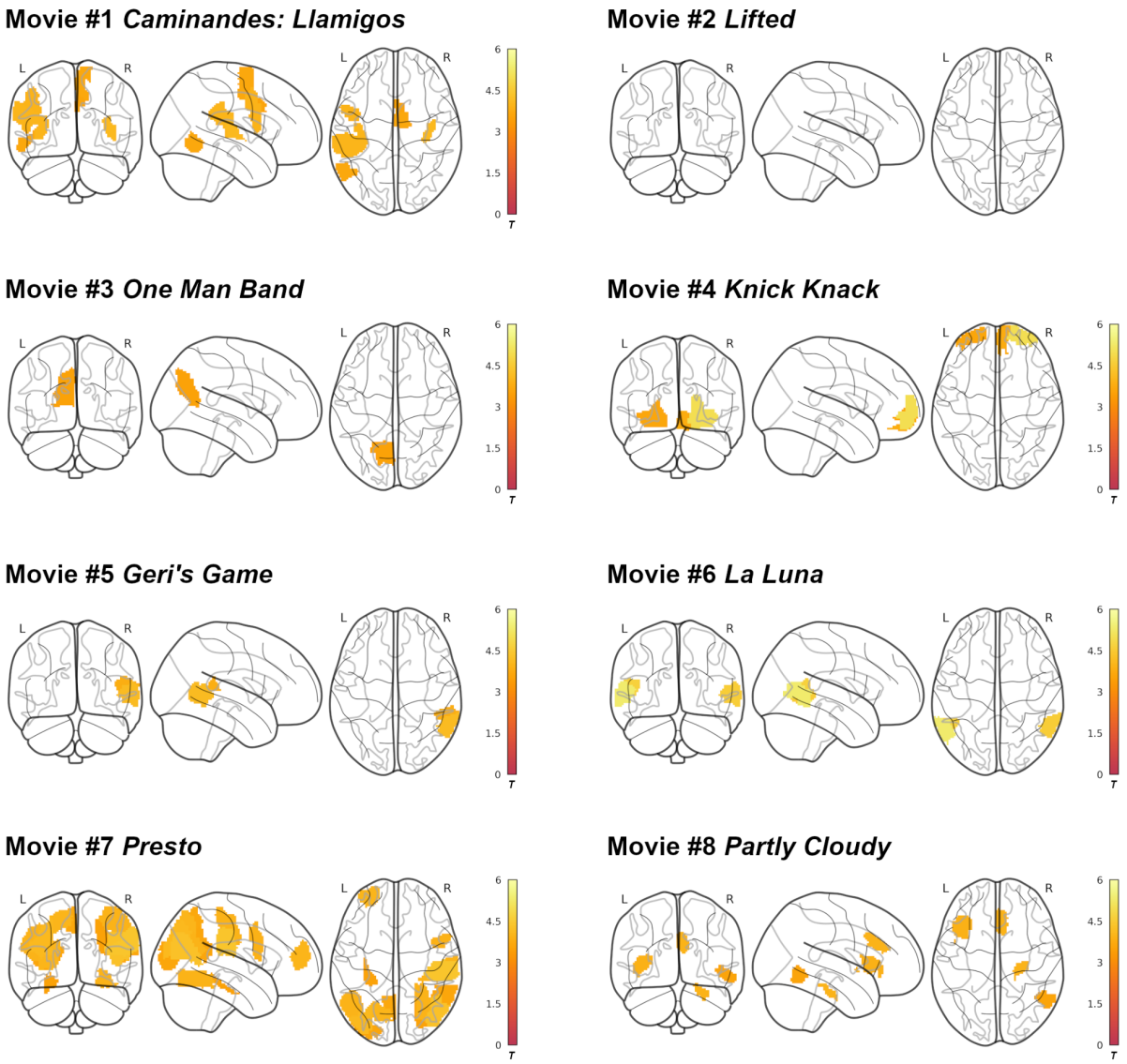
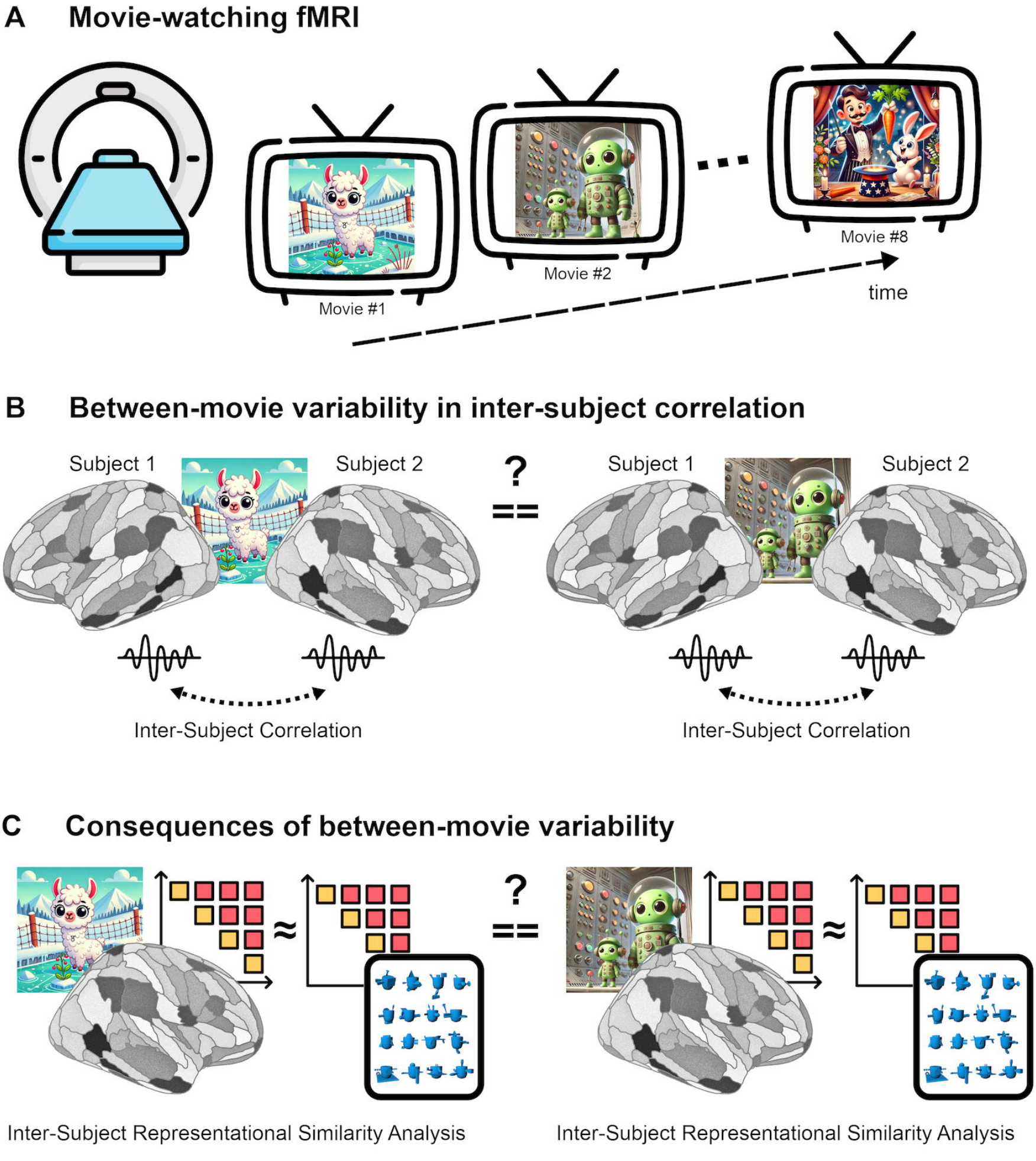
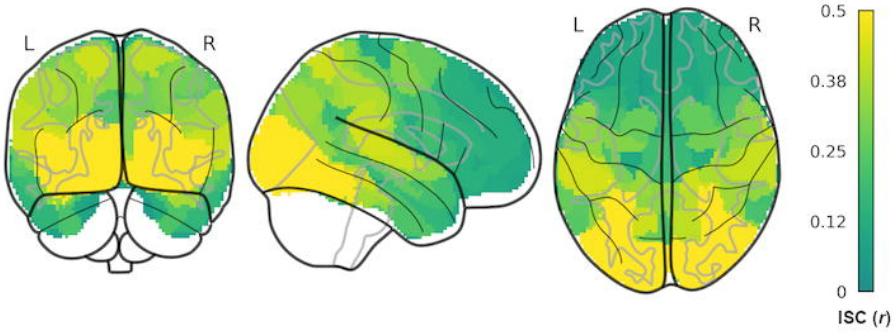


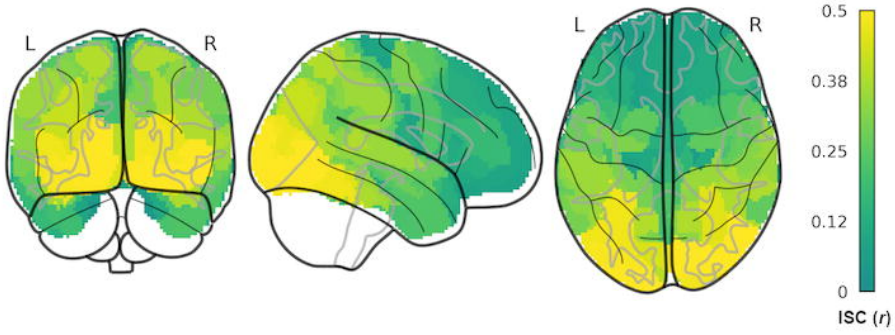
Figure 4. Between-movie ISC variability leads to varying IS-RSA associations between a state-like behavior and neural synchronization.
We used inter-subject representational similarity analysis (IS-RSA) to discover brain regions in which subject similarity in conceptual representations of novel objects are associated with subject similarity in neural synchronization during movie watching. The results of the IS-RSA revealed associations between semantic representations of novel objects and ISC in distinct brain regions for each movie. Thus, between-movie variability in ISC has detrimental downstream effects for generalizability of findings with a single movie clip as these effects likely do not generalize to other movies.


A
Movie-watching fMRI
Movie #1
Movie #2
Movie #8
time
B
Between-movie variability in inter-subject correlation
Subject 1
Subject 2
?
==
Subject 1
Subject 2
Inter-Subject Correlation
Inter-Subject Correlation
C
Consequences of between-movie variability
≈
?
==
≈
Inter-Subject Representational Similarity Analysis
Inter-Subject Representational Similarity Analysis

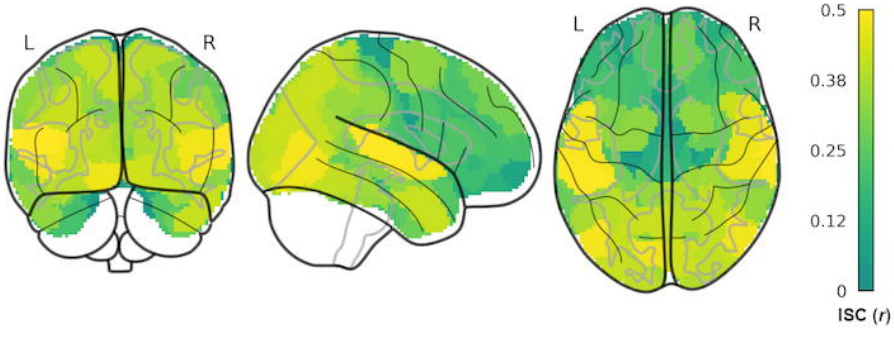
Movie #1 *Caminandes: Llamigos*



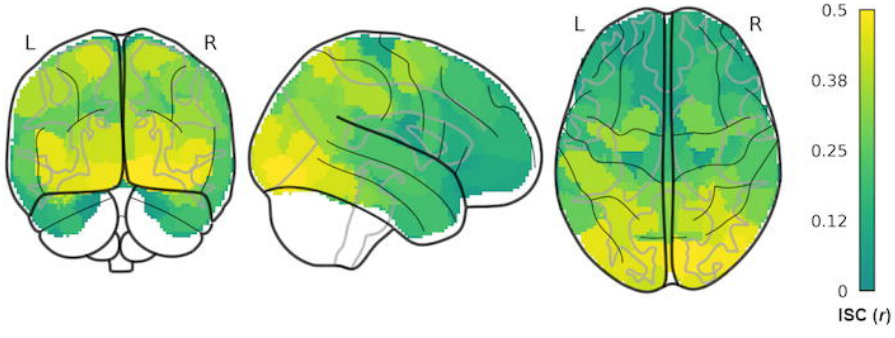
Movie #2 *Lifted*



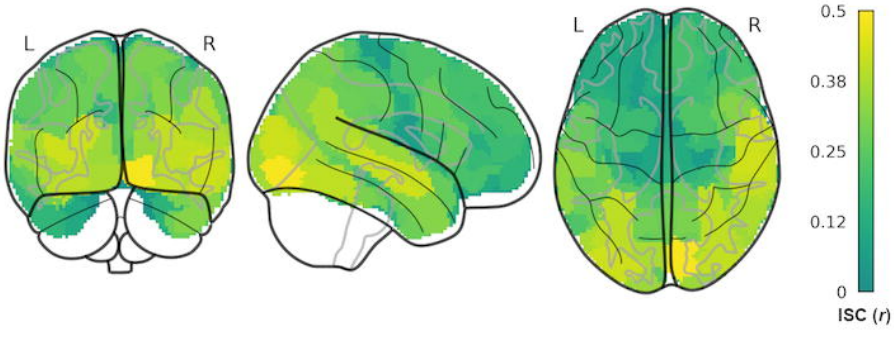
Movie #3 *One Man Band*



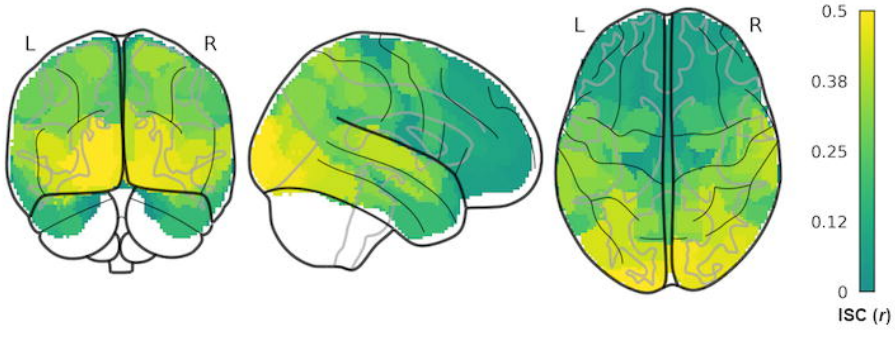
Movie #4 *Knick Knack*



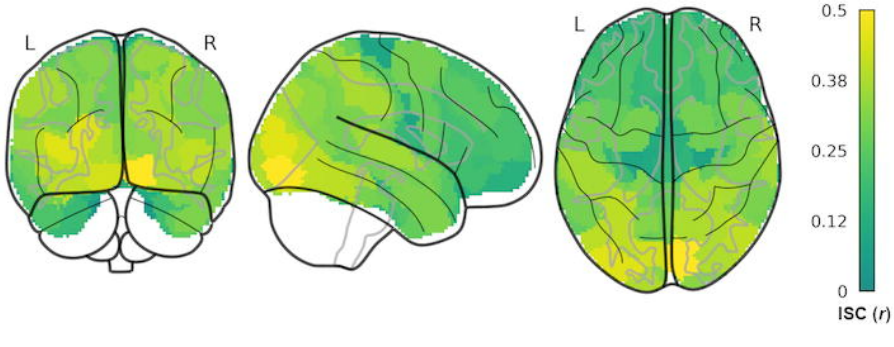
Movie #5 *Geri's Game*



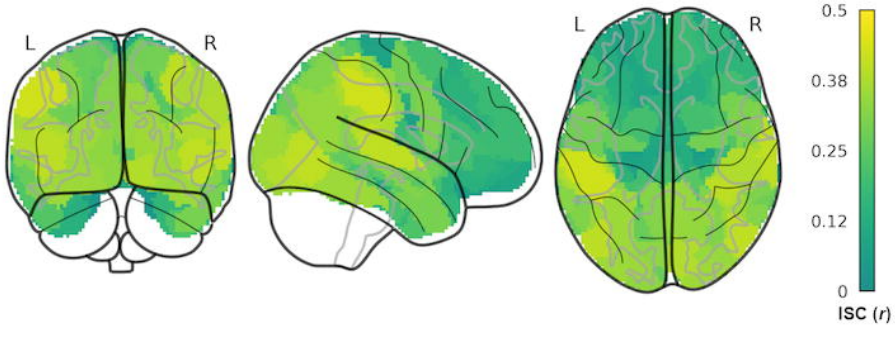
Movie #6 *La Luna*

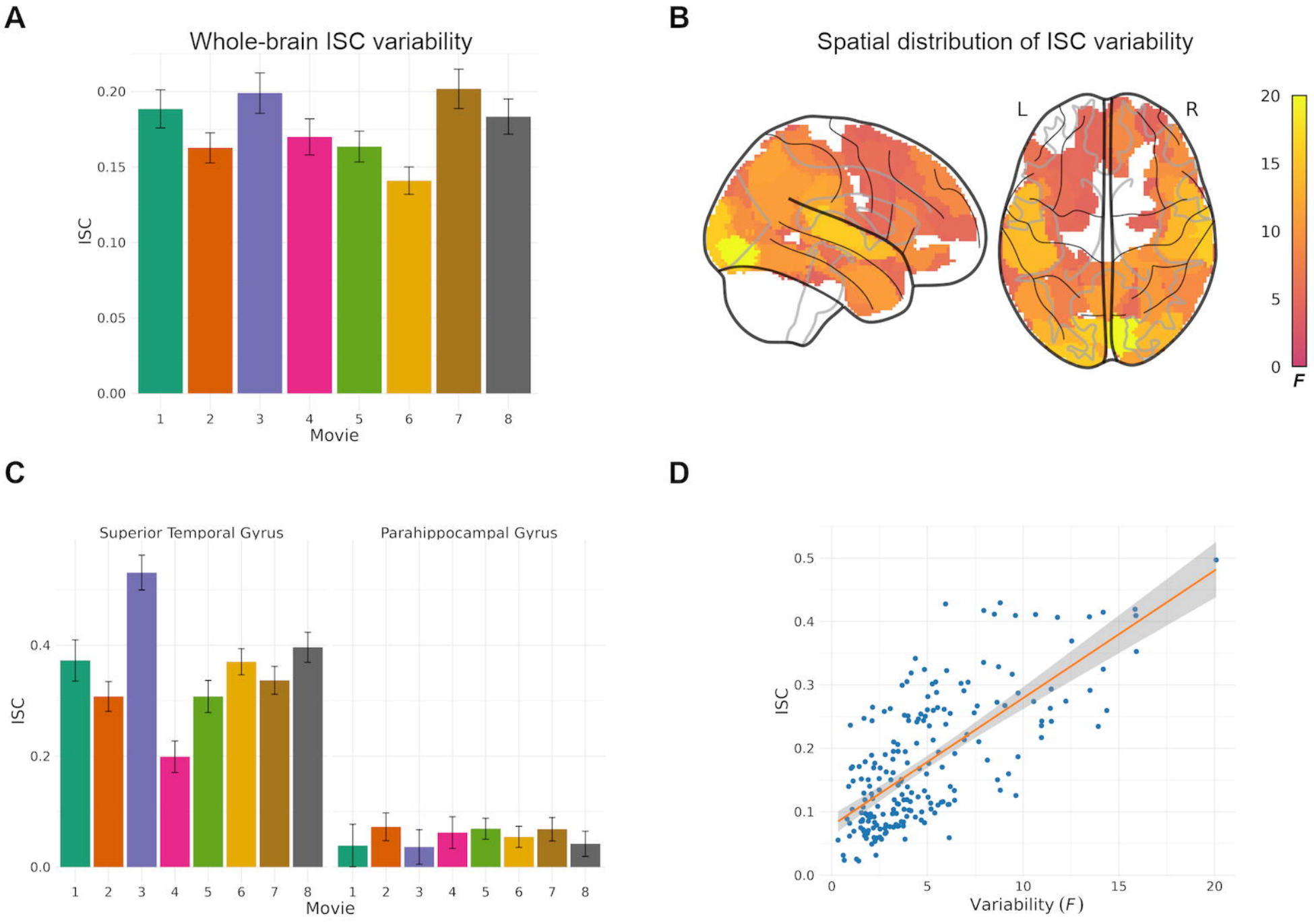


Movie #7 *Presto*

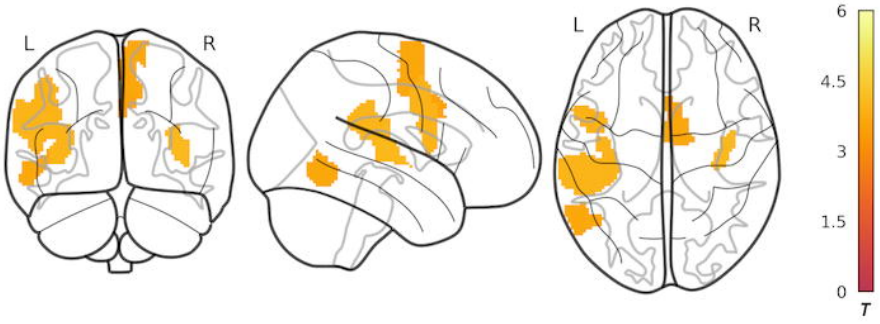


Movie #8 *Partly Cloudy*

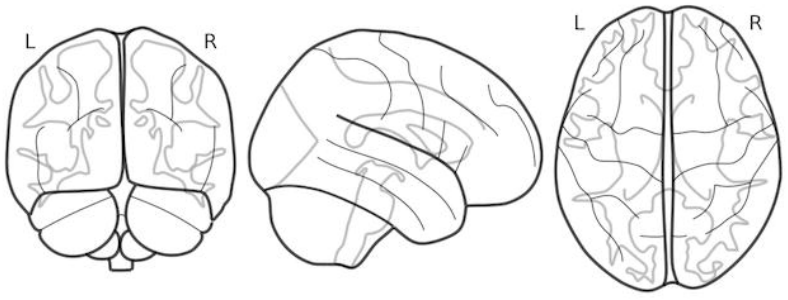



A
Whole-brain ISC variability
ISC
0.00
0.05
0.10
0.15
0.20
1
2
3
4
5
6
7
8
Movie
B
Spatial distribution of ISC variability
L
R
20
15
10
5
0
F
C
Superior Temporal Gyrus
Parahippocampal Gyrus
ISC
0.0
0.2
0.4
1
2
3
4
5
6
7
8
Movie
D
ISC
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
Variability (F)

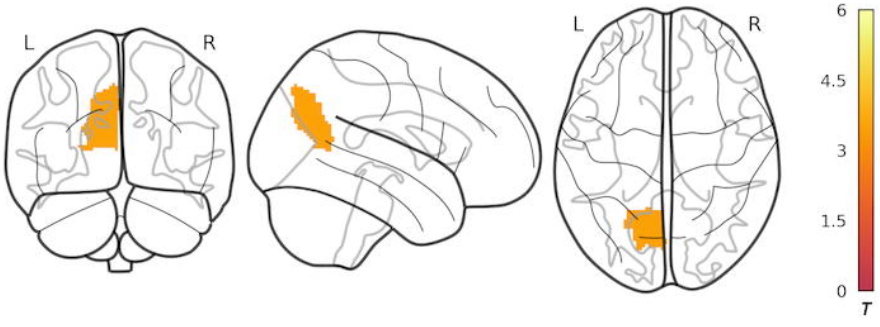
Movie #1 *Caminandes: Llamigos*



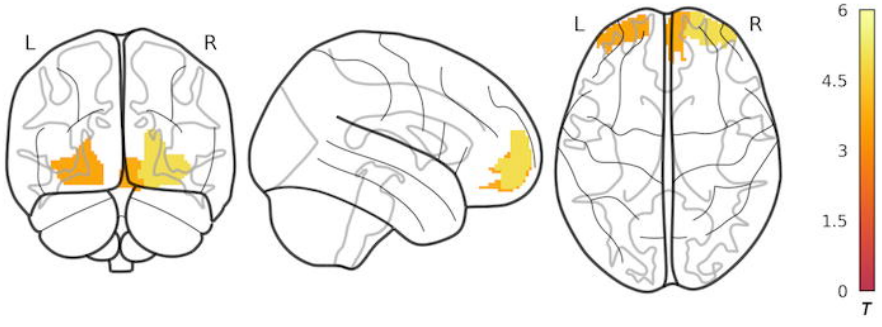
Movie #2 *Lifted*



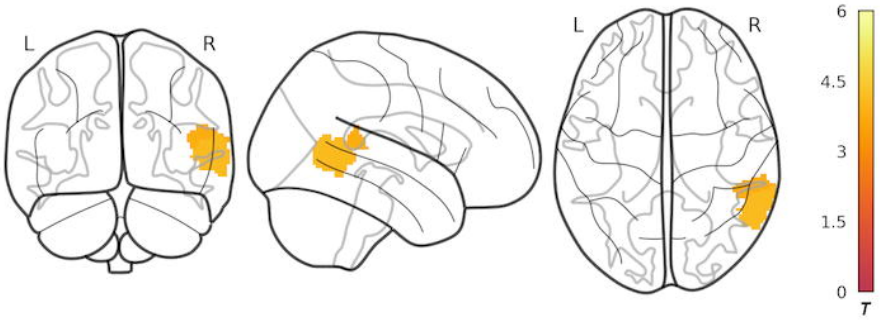
Movie #3 *One Man Band*



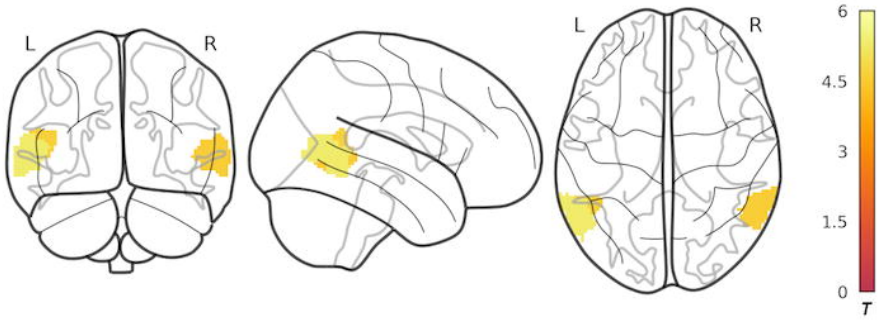
Movie #4 *Knick Knack*



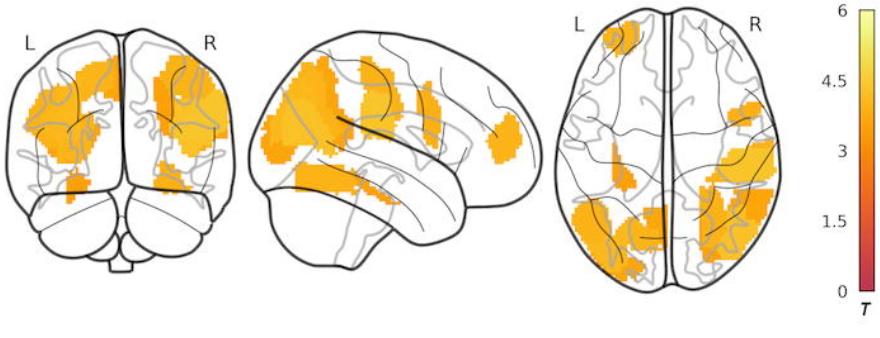
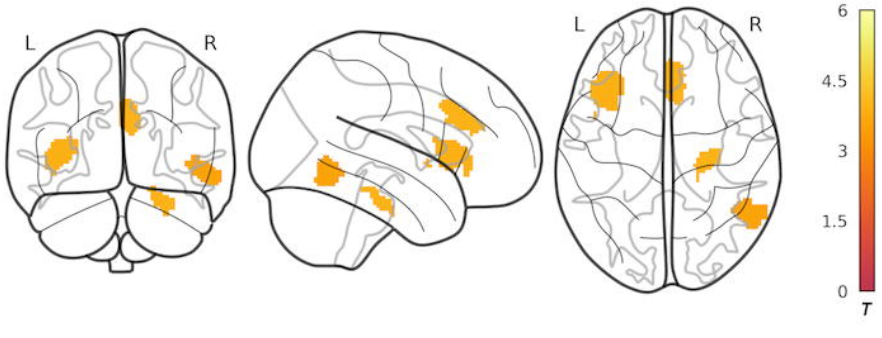
Movie #5 *Geri's Game*



Movie #6 *La Luna*



Movie #7 *Presto*



Movie #8 *Partly Cloudy*