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] Social Brain Sciences Lab, 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” neuroimaging paradigms, where subjects watch movies during fMRI, have become increasingly popular. Movie fMRI data is often analyzed using inter-subject correlation (ISC), which measures the similarity of neural time series across individuals. Differences in ISC during movie-watching have been linked to individual traits and states. However, movies are complex cultural artifacts that differ in content, structure, and style, raising concerns about the generalizability of ISC findings across movies.

Using fMRI data from 112 subjects watching eight animated movies, we quantified between-movie variability in ISC and examined its implications for trait- or state-like effects. ISC varied substantially across movies and brain regions, with the strongest ISC observed in regions showing the greatest variability. Consequently, associations between behavior and ISC differed markedly across movies.

These findings suggest that movie-based paradigms should be treated like task-based designs, requiring clear specification of movie features. Broad claims about “movie watching” are not warranted.

Introduction

In the last two decades, “naturalistic” paradigms have gained enormous popularity in neuroscience [1–6]. These paradigms include subjects viewing movies [7,8] or listening to narratives [9–11] or music [12,13] while undergoing functional brain imaging, most commonly functional magnetic resonance imaging (fMRI). Although the term “naturalistic” is used variably across subfields [14], here we use it specifically to refer to the use of media materials such as movies or audiobooks in neuroimaging research. It has been argued that naturalistic paradigms more closely resemble everyday life experiences [5,15], thereby addressing the limited “ecological validity” [16–18] of traditional, cognitively controlled neuroscience experiments. Specifically, the simple and arbitrarily ordered stimuli used in conventional experiments might restrict the generalizability of findings to everyday-life settings [15]. However, to date, the generalizability of findings from naturalistic paradigms has been largely overlooked, despite their reliance on culturally contingent and idiosyncratic artifacts such as movies and literary narratives, which are forms of media deliberately designed to serve specific purposes, such as entertaining or informing audiences [14,19].

A 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 [20]: Calculating the correlation coefficient of the two time series yields a measure of synchronization of brain activation between the different subjects [21]. Summary statistics, such as the average correlation between all possible pairs of subjects, are calculated to express the level of synchronization across all subjects 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 [7,21]. 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 subjects remain largely uncorrelated and do not contribute to the ISC [20,21]. 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” [22], or having similar beliefs about the characters in a story [23], 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 [24]. Friends show higher ISC than individuals who are further away from each other in a social network [25]. Sisters show higher ISC than friends [26], and married couples show higher ISC than random couples [27]. In the clinical domain, alterations in ISC during movie watching have been reported in individuals with autism [28–30], psychosis [31], depression [32], and social anxiety [33], 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 [34,35]. For example, subjects who are similar in age [36–38], personality traits [39,40], political orientation [41–44], sexual preferences [45], or reading ability [46] 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 trait- or state-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 [24,27,33,44]. 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 [7,47]. Different movies vary along multiple features. Considering that the main

measure of neural synchronization, ISC, is strongly stimulus-driven [20,21], different movies might lead to different trait- or state-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 [48], as feature films evoke much stronger ISC than filmed segments of reality [48] or educational movies [38].

Surprisingly, this systematic variation in ISC between movies has been neglected when considering trait- and state-like 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 trait- or state-like differences in ISC are bound to the movie being shown?

Here, we leverage a dataset of 112 adults watching eight movies each [49] to (i) quantify between-movie variability in ISC across the brain and (ii) assess the potential consequences of this variability in ISC for trait- or state-like differences in ISC (**Figure 1**). Specifically, we assessed how between-movie variability affects association between a cognitive 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 [49]. 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 [50]. 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 ISC patterns appeared qualitatively similar across movies. In prior work, such coarse similarities have been interpreted as evidence that different naturalistic materials evoke broadly comparable synchronization across the brain [51]. However, closer visual inspection of the ISC maps in **Figure 2** suggests that different movies elicited distinct neural synchronization patterns. For example, Movie #3 appeared to elicit generally higher ISC values than Movie #5.

Building on these qualitative observations, we next quantified the variability in ISC between movies through statistical analysis. 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 [20]. Then, we submitted 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 ($F(7,385) = 4.93$, $p < 0.001$, generalized eta-squared [η^2_G] = 0.051), indicating considerable 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 97 of the 210 cortical parcels (46.2%). 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.

To further test whether ISC patterns carry movie-specific information, we ran a cross-validated multivariate classification analysis to distinguish the movies based on parcel-wise ISC patterns, achieving an accuracy of 62.9% (chance level = 12.5%; $p = 0.001$, permutation test). This demonstrates that the spatial distribution of ISC values contains sufficient information to robustly differentiate between movies. The analysis is described in detail in the **Supplementary Results** and illustrated in **Supplementary Figure 4**.

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 ($r = 0.73$, $t(208) = 15.62$, $p < 0.001$, $R^2 = 0.54$). While this relationship is not surprising from a statistical standpoint, since regions with higher ISC allow for more variability, it is still important to highlight. It indicates that brain regions with stronger ISC also show greater differences in ISC between movies. In other words, the regions that most consistently synchronize across individuals are also the ones most affected by between-movie variability.

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

Our next objective was to assess consequences of between-movie variability in ISC for the generalizability of trait- or state-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* [49,52]. 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 [23]. 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 [53]. 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 [49], 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 subjects 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 [54] 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 trait- or state-like differences in ISC

[37,39,41,44,46,55,56]. 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 trait- or state-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, we found that similarities between subjects in semantic representations were statistically significantly associated with neural synchronization between subjects in parcels within bilateral insula (right hemisphere: $p_{FWE} = 0.004$ and left hemisphere: $p_{FWE} = 0.007$), a parcel in the left inferior parietal lobule ($p_{FWE} = 0.006$), the left precentral gyrus ($p_{FWE} = 0.007$), the right superior frontal gyrus ($p_{FWE} = 0.02$), the left inferior temporal gyrus ($p_{FWE} = 0.03$), and the right anterior cingulate ($p_{FWE} = 0.03$). In contrast, for Movie #2, we found no statistically significant association between semantic representations and neural synchronization between subjects. For Movie #3, we found significant associations of similarities in semantic representations with ISC in the left precuneus ($p_{FWE} < 0.05$), which did not overlap with any of the regions that were found for Movie #1. For Movie #4, we found significant associations in bilateral middle frontal gyrus (right hemisphere: $p_{FWE} < 0.001$ and left hemisphere: $p_{FWE} = 0.03$) and the right orbital gyrus ($p_{FWE} = 0.02$), which again did not overlap with regions found in Movie #1 or #3. Likewise, for Movie #5 through Movie #8, we found non-overlapping brain regions that showed an association between semantic representations and neural synchronization except for a single parcel in the right middle temporal gyrus, which showed an association both for Movie #5 ($p_{FWE} = 0.004$) and Movie #6 ($p_{FWE} < 0.001$). Details on the statistically significant IS-RSA associations for all movies are listed in **Supplementary Table 3**. Details on all parcels and movies are available online (https://dx.doi.org/10.17605/OSF.IO/S78WU). Thus, between-movie variability in ISC has detrimental downstream effects for generalizability of findings with one movie as these effects likely do not generalize to other movies.

A subset of brain regions show IS-RSA effects that generalize across movies

To examine whether any associations generalized across movies, we conducted an additional IS-RSA that included a random effect for movie. This analysis identified three brain regions, including bilateral posterior inferior temporal gyrus and right posterior superior temporal sulcus, where ISC was significantly associated with semantic representations of novel objects across movies (see **Supplementary Results**).

Discussion

Here, we quantified the extent of variability in ISC between different movies, and we assessed the consequences of this variability for investigations of trait-like or state-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 trait-like and state-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 trait-like and state-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 [7,47], it has been clear that movies are a highly complex and multidimensional sequence of stimuli, making it challenging to isolate specific features that drive neural synchronization [2,7]. This is further complicated by the fact that many features in movies are correlated, making it difficult to isolate their individual contributions [57]. 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 the limited generalizability of findings based on a single or small set of movies is to include movie as a "random effect" in the statistical model. Following Yarkoni [58], "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." [58 p. 3]. With rare exceptions [44], studies investigating trait- or state-like differences in ISC that include multiple movie clips have either modeled the stimulus as a fixed effect [24] or run separate models per movie, effectively treating the stimulus as fixed. This limits the ability to generalize beyond the specific stimulus set used, a problem long recognized in neuroscience [59] and other fields [58,60,61]. It is therefore possible that the phenomenon we describe here reflects a specific instance of the "stimulus-as-fixed-effect fallacy" which, in principle, can be addressed by including the movie as a random effect in the statistical model. To explore this approach, we tested an IS-RSA model that included movie as a random effect. This analysis revealed consistent associations in a small number of brain regions, but most effects remained movie specific.

These findings raise a deeper issue: what exactly are we generalizing across? Can we reasonably assume that all movies sample from the same underlying population, given their substantial variability in structure, content, and style? Do different movies address the same research question, as the random-effects framework assumes, or do they address distinct ones? The central challenge persists: even when between-movie variability is statistically modeled, we lack a principled understanding of which features within these complex, multidimensional sequences of stimuli give rise to ISC and IS-RSA effects. Without identifying and quantifying the relevant features, generalization remains difficult.

Movies are more akin to tasks than to rest

Naturalistic paradigms have been proposed as an alternative to resting-state paradigms [5], and movie watching addresses several limitations of resting-state acquisitions. For example, ISC during movie viewing shows higher test-retest reliability when the same movie is watched multiple times [62]. Movies are also well-suited for pediatric populations [6], as they enhance attentional engagement and reduce head motion compared to resting-state scans [36,63].

However, our findings suggest that different movies elicit distinct neural synchronization patterns across individuals, making them more comparable to separate tasks than to resting-state conditions [14]. Consequently, studies using movie paradigms should document stimulus properties with the same level of rigor applied in task-based research, even though the relevant features may be difficult to isolate [6]. Researchers should explicitly specify which perceptual, cognitive, or affective processes a given movie is intended to engage. Machine learning-based annotation tools, such as those implemented in *pliers* [64] or *NeuroScout* [65], can help address this

291 by automatically extracting auditory, visual, linguistic, or affective features from naturalistic
292 materials. These tools provide a data-driven foundation for identifying features that may
293 contribute to ISC and its variability across movies, offering a promising path toward greater
294 interpretability and generalizability of ISC-based findings. Recent work also highlights the
295 potential of combining stimulus-based annotations with multi-dimensional experience sampling,
296 which captures subjects' ongoing thought patterns without disrupting the movie-watching
297 experience [66]. Combining such subjective reports with automated feature annotations could
298 clarify how both stimulus features and internal cognitive states shape ISC.

299 By contrast, when the goal is to evoke states closer to rest, researchers may turn to specially
300 designed audiovisual materials such as *Inscapes*, which have been shown to elicit more rest-
301 like brain activation patterns than conventional movie clips [63].

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

303 It has been argued that the rich and complex stimulation provided by movies more closely
304 resembles everyday-life perception than the custom-designed stimuli typically used in laboratory
305 settings [7]. In a seminal study where subjects watched a James Bond movie, the paradigm was
306 described as observing a "dynamic natural scene" [47]. However, a Hollywood-produced action
307 movie is far from natural. At best, movies are abstracted representations of daily-life
308 experiences. They contain elements deliberately manipulated by filmmakers to capture and
309 maintain the audience's attention, such as camera movements, zooms, editing techniques like
310 cuts, and soundtracks. Indeed, researchers previously noted that critically acclaimed movies
311 evoke much higher ISC than actual natural scenes [48,67] or educational films [38]. While naturalistic
312 paradigms may appear more "ecologically valid" than traditional controlled experiments, it
313 remains unclear to what extent ISC during movie watching is driven by stylistic, artificial features
314 such as cuts and zooms [68–70]. Variation in these features across films is directly relevant to the
315 issue of between-movie variability, as they may contribute to differences in neural
316 synchronization.

317 **Rethinking the use of "naturalistic" in movie fMRI research**

318 The term "naturalistic" is widely used in neuroscience to describe paradigms intended to
319 approximate everyday-life experiences [5]. However, the term is highly ambiguous [14]. In vision
320 science, "naturalistic" might refer to static images of everyday-life objects or scenes [71–73], while
321 in social neuroscience it may describe mobile neuroimaging conducted in everyday-life settings

[74,75]. In the context of fMRI, the term is often applied to studies using movies or narratives because such media are part of everyday life for many individuals. Yet, as others have argued [14,19], movies are deliberately constructed artifacts, designed to entertain, inform, or persuade. Labeling them as “naturalistic” risks obscuring their inherently artificial, highly structured, and culturally specific nature. Rather than relying on this broad and inconsistently applied term, we suggest that researchers describe movie stimuli in terms of their specific properties, such as being dynamic, high-dimensional, and temporally extended, and explicitly articulate how these properties relate to the perceptual, cognitive, or behavioral phenomena under investigation [cf. 76].

Constraints on generality, limitations, and future directions

It remains an open question to what extent our findings generalize across different uses of naturalistic materials in neuroscience. For example, studies employing control conditions such as scrambled movies [8,9] or scrambled music [12] may be less affected by the stimulus-specificity in ISC observed here. The rationale behind scrambled conditions is that many (though likely not all) features of naturalistic materials that drive variability in ISC are held constant between the intact and scrambled versions, isolating only those of interest to the researcher. However, if this assumption is not fully met, for example 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. Our findings may also be less consequential for studies focusing on specific types of events embedded within naturalistic materials, such as those investigating memory or event segmentation [77–79], many of which use the general linear model to examine brain responses to specific event types [47,80–82]. In such cases, variability between entire movies may have minimal impact on the targeted event-level effects. Our findings are likely not relevant for studies using voxel-wise encoding models to map neural responses to annotated stimulus features, as these models are typically evaluated on independent datasets, providing a direct assessment of generalizability across different stimuli [83,84].

Within-subject functional connectivity measures may be more consistent across different movies than ISC-based measures [85]. This relative stability is likely because functional connectivity is largely driven by intrinsic network architecture [86], which is commonly measured during rest and shows only modest modulation by movies [63] or task conditions [87,88]. However, movie-specific differences in functional connectivity may still be meaningful in certain contexts [89,90], such as when using functional connectivity patterns to predict individual phenotypes [91].

More broadly, the adverse effects of between-stimulus variability on the generalizability of trait- and state-like differences in ISC are unlikely to be limited to movie fMRI. They may extend to other classes of naturalistic materials, such as narratives or music; to analytical approaches such as inter-subject functional correlation [cf. 92] and ISC-based predictive modeling of individual phenotypes [cf. 67]; and to neuroimaging modalities beyond fMRI, including magnetoencephalography (MEG), electroencephalography (EEG) [cf. 38], or functional near-infrared spectroscopy (fNIRS). Most of these possibilities remain to be tested and require systematic empirical investigation in future research.

A limitation of the current study is that the dataset was not designed to test whether more natural stimuli, such as unedited natural scenes, evoke less variable ISC than feature films. Prior work suggests that natural scenes elicit substantially lower ISC than Hollywood movies [48,67]. In our data, brain regions with lower ISC also show less between-movie variability (**Figure 2D**), raising the possibility that more natural scenes may yield reduced ISC variability. This remains an important question for future research to address.

A further limitation is that our study did not include direct measures or manipulations of subjects' psychological states during movie viewing. While the semantic representations assessed here may contain state-like components [49], we cannot directly assess transient state-like effects or their relationship with ISC. Future work should aim to manipulate such states more directly and test whether their associations with ISC generalize across different movies.

Future work could also compare variability in responses to movies or narratives with variability across cognitive tasks that target similar underlying processes [93–95]. Such comparisons may help contextualize movie-induced variability relative to more traditional task-based paradigms. Making such a comparison meaningful would require identifying the specific perceptual or cognitive processes that a given set of naturalistic materials is likely to engage. More generally, if the goal is to identify common neural patterns that emerge across different movies during free viewing, data-driven approaches such as factor analysis may be useful for isolating components that reflect shared structure beyond movie-specific effects.

Conclusion

We find substantial between-movie variability in ISC across the cortex, leading to distinct spatial distributions of trait- or state-like differences in the brain for separate movies, with minimal overlap between movies. Consequently, findings of trait-like or state-like differences in ISC

based on a single movie are not generalizable across different movies, raising questions about 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 [49]. These subjects were the same as those included in the fMRI analysis in the dataset description paper by Eijk et al. [49] 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) [96] 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,

415 Germany). Blood oxygenation level dependent (BOLD) functional images were acquired using a
416 multi-band 2D-echo planar imaging (EPI) sequence (Human Connectome Project) [97] with
417 repetition time (TR) = 1000 ms, echo time (TE) = 34 ms, flip angle = 60º, 2mm isotropic
418 resolution, field of view (FOV) = 208x208x132 mm. A multi-band acceleration factor of 6 was
419 used in the slice direction, and no parallel imaging was applied in-plane. Phase encoding was
420 applied in the anterior-posterior (AP) direction with a partial Fourier coverage of 7/8. About 2074
421 volumes were acquired during movie watching. Anatomical images were acquired with a T1-
422 weighted 3D MPRAGE sequence, integrated parallel acquisition technique (iPAT) acceleration
423 factor = 2, with TR = 2200 ms, TE = 2 ms, inversion time (TI) = 1100 ms, flip angle = 11°, 0.8
424 mm isotropic resolution, FOV = 256x240x166 mm; and with a T2-weighted turbo spin echo
425 (TSE) sequence with variable flip angle, TR = 3200 ms, TE = 569 ms, echo spacing = 4.12 ms,
426 turbo factor = 314, 0.8 mm isotropic resolution, and FOV = 256x240x166 mm. Raw imaging
427 data were converted to Brain Imaging Data Structure (BIDS) format [98] using *BIDScoin* (version
428 1.5; RRID:SCR_022839) [99].

429 **Imaging data preprocessing**

430 Each subject's raw imaging data underwent a quality control assessment for excessive motion,
431 susceptibility artifacts, or other sources of noise using *MRIQC* (version 23.1.0;
432 RRID:SCR_022942). Imaging data preprocessing was implemented using *fMRIPrep* (version
433 20.2.7; RRID:SCR_016216) [100], which is based on *Nipype* (version 1.7.0; RRID:SCR_002502)
434 [101]. The main preprocessing steps included segmentation, surface reconstruction, and spatial
435 normalization of the T1-weighted anatomical image; and slice timing correction, susceptibility
436 distortion correction, head-motion correction, and spatial normalization of the fMRI data. Details
437 of the preprocessing pipeline are included in the **Supplementary Methods**.

438 **fMRI data denoising**

439 Each subject's preprocessed fMRI data underwent further denoising, implemented using *Nilearn*
440 (version 0.9.2; RRID:SCR_001362) in *Python* (version 3.7.13; RRID:SCR_008394). The voxel-
441 wise fMRI time series were detrended and low-pass filtered at 0.08 Hz. Next, the following
442 confounds estimated by fMRIprep during preprocessing were regressed out from the voxel-wise
443 time series: six motion parameters (three translations, three rotations) and their first temporal
444 derivatives, as well as cerebrospinal fluid and white matter time series and their first temporal
445 derivatives. In addition, scans that exceeded a threshold of 0.5 mm framewise displacement

446 (FD) or 1.5 standardized DVARS were annotated as motion outliers and included as nuisance
447 regressors. Subsequently, the time series were standardized to have a mean of zero and unit
448 variance. After denoising, we segmented the fMRI time series and extracted a separate time
449 series for each movie clip based on the onset and offset times recorded in the Presentation log
450 files. The denoised fMRI data for each movie was then spatially smoothed with a 6 mm full-
451 width-at-half-maximum three-dimensional Gaussian kernel.

452 **Inter-subject correlation (ISC) analysis**

453 To calculate parcel-wise ISC values, we first extracted a spatially averaged parcel-wise time
454 series for each subject and movie from each of the 210 cortical parcels defined by the
455 Brainnetome atlas [50] using *NLtools* (version 0.4.7) [102] in *Python*. For each movie and cortical
456 parcel, we then calculated the ISC values for all possible subject pairs by computing the
457 Pearson correlation coefficient between the parcel-wise fMRI time series of subject [i] and
458 subject [j] of each pair [i,j] using *scikit-learn* (version 1.0.2; RRID:SCR_002577) in *Python*. To
459 obtain a mean ISC value for each parcel, we averaged the ISC values across all subject pairs.
460 Visualization of the ISC values in the brain was implemented using *Nilearn*.

461 **Statistical analyses of ISC values**

462 Prior to statistical analysis, all ISC values were transformed using Fisher's r-to-z transformation
463 to normalize the distribution of correlation coefficients. To quantify between-movie variability in
464 ISC across the whole brain, we averaged the ISC values across all cortical parcels to obtain a
465 single ISC value per subject pair and movie. Then, we entered the whole-brain ISC values into a
466 repeated-measures ANOVA with Movie as the within-subject factor. We used data from 56
467 subject pairs, ensuring that each subject contributed only to a single ISC value per movie and
468 that the assumptions of the repeated-measures ANOVA were not violated. Next, we repeated
469 the ANOVA for each parcel separately to obtain a spatial distribution of between-movie
470 variability across the brain. The ANOVAs were performed using the *ez* (version 4.4-0;
471 RRID:SCR_020990) and *rstatix* (version 0.7.2; RRID:SCR_021240) packages in *R*.

472 To investigate whether brain regions with high levels of ISC also show high variability in ISC, we
473 correlated the parcel-wise level of ISC with the parcel-wise between-movie variability in ISC.
474 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 [34,35] using linear mixed-effects models [44,54] 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**.

Control analyses

We ran multiple additional analyses to test the robustness of our findings, which are detailed in the **Supplementary Results**. These include: (1) re-running all ISC and IS-RSA analyses using truncated time series to control for differences in movie length; (2) implementing a resampling approach to assess the stability of between-movie variability in ISC across different non-overlapping subject-pair configurations; (3) incorporating a random effect for movie in the IS-RSA model to evaluate whether any effects generalize across our set of movies; (4) re-running IS-RSA analyses without covariates; and (5) repeating IS-RSA analyses using an alternative distance measure to quantify subject similarity in conceptual representations.

Code availability

Python and R code is available online (https://doi.org/10.5281/zenodo.15929487).

Data availability

Raw data is available in the Radboud Data Repository, as described in the accompanying data descriptor (https://doi.org/10.1016/j.neuroimage.2022.119734).

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. Grall, C. & Finn, E. S. Leveraging the power of media to drive cognition: a media-informed approach to naturalistic neuroscience. *Soc. Cogn. Affect. Neurosci.* **17**, 598–608 (2022).

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

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

17. Holleman, G. A., Hooge, I. T. C., Kemner, C. & Hessels, R. S. The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity. *Front. Psychol.* **11**, (2020).

18. Kihlstrom, J. F. Ecological Validity and "Ecological Validity". *Perspect. Psychol. Sci.* **16**, 466–471 (2021).

19. Schmälzle, R. & Huskey, R. Integrating media content analysis, reception analysis, and media effects studies. *Front. Neurosci.* **17**, (2023).

20. 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).

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

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

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

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

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

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

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

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

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

30. Mangnus, M. *et al.* Preserved Spontaneous Mentalizing Amid Reduced Intersubject Variability in Autism During a Movie Narrative. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* (2024) doi:10.1016/j.bpsc.2024.10.007.

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

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

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

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

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

36. 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).

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

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

39. 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).

40. 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).

41. 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).

42. 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).

43. 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).

44. 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).

45. 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).

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

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

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

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

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

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

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

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

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

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

627 56. Baek, E. C. *et al.* Lonely Individuals Process the World in Idiosyncratic Ways. *Psychol. Sci.*
628 **34**, 683–695 (2023).

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

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

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

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

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

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

638 63. Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C. & Castellanos, F. X. Inscapes: A movie
639 paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*
640 **122**, 222–232 (2015).

641 64. McNamara, Q., De La Vega, A. & Yarkoni, T. Developing a Comprehensive Framework for
642 Multimodal Feature Extraction. in *Proceedings of the 23rd ACM SIGKDD International*
643 *Conference on Knowledge Discovery and Data Mining* 1567–1574 (Association for
644 Computing Machinery, New York, NY, USA, 2017). doi:10.1145/3097983.3098075.

645 65. de la Vega, A. *et al.* Neuroscout, a unified platform for generalizable and reproducible fMRI
646 research. *eLife* **11**, e79277 (2022).

647 66. Wallace, R. S. *et al.* Mapping patterns of thought onto brain activity during movie-watching.
648 *eLife* **13**, RP97731 (2025).

649 67. Li, X., Eickhoff, S. B. & Weis, S. Stimulus Selection Influences Prediction of Individual
650 Phenotypes in Naturalistic Conditions. *Hum. Brain Mapp.* **46**, e70164 (2025).

651 68. Herbec, A., Kauppi, J.-P., Jola, C., Tohka, J. & Pollick, F. E. Differences in fMRI intersubject
652 correlation while viewing unedited and edited videos of dance performance. *Cortex* **71**, 341–
653 348 (2015).

654 69. Heimann, K. S. *et al.* “Cuts in Action”: A High-Density EEG Study Investigating the Neural
655 Correlates of Different Editing Techniques in Film. *Cogn. Sci.* **41**, 1555–1588 (2017).

70. Smith, T. J., Levin, D. & Cutting, J. E. A Window on Reality: Perceiving Edited Moving Images. *Curr. Dir. Psychol. Sci.* **21**, 107–113 (2012).

71. Allen, E. J. *et al.* A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* **25**, 116–126 (2022).

72. Hebart, M. N. *et al.* THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE* **14**, e0223792 (2019).

73. Hebart, M. N. *et al.* THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *eLife* **12**, e82580 (2023).

74. Vigliocco, G. *et al.* Ecological brain: reframing the study of human behaviour and cognition. *R. Soc. Open Sci.* **11**, 240762 (2024).

75. Shamay-Tsoory, S. G. & Mendelsohn, A. Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspect. Psychol. Sci.* **14**, 841–859 (2019).

76. Holleman, G. A., Hooge, I. T. C., Kemner, C. & Hessels, R. S. The Reality of “Real-Life” Neuroscience: A Commentary on Shamay-Tsoory and Mendelsohn (2019). *Perspect. Psychol. Sci.* **16**, 461–465 (2021).

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

78. Ben-Yakov, A. & Henson, R. N. The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *J. Neurosci.* **38**, 10057–10068 (2018).

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

80. 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).

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

82. 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).

83. 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).

84. Dupré la Tour, T., Visconti di Oleggio Castello, M. & Gallant, J. L. The Voxelwise Encoding Model framework: A tutorial introduction to fitting encoding models to fMRI data. *Imaging Neurosci.* **3**, imag_a_00575 (2025).

85. Tian, L., Ye, M., Chen, C., Cao, X. & Shen, T. Consistency of functional connectivity across different movies. *NeuroImage* **233**, 117926 (2021).

86. Buckner, R. L., Krienen, F. M. & Yeo, B. T. T. Opportunities and limitations of intrinsic functional connectivity MRI. *Nat. Neurosci.* **16**, 832–837 (2013).

87. Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S. & Petersen, S. E. Intrinsic and Task-Evoked Network Architectures of the Human Brain. *Neuron* **83**, 238–251 (2014).

88. 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).

89. Vanderwal, T. *et al.* Individual differences in functional connectivity during naturalistic viewing conditions. *NeuroImage* **157**, 521–530 (2017).

90. Kröll, J.-P. *et al.* Naturalistic viewing increases individual identifiability based on connectivity within functional brain networks. *NeuroImage* **273**, 120083 (2023).

91. Finn, E. S. & Bandettini, P. A. Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage* **235**, 117963 (2021).

92. 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).

93. Andermane, N., Bosten, J. M., Seth, A. K. & Ward, J. Individual differences in change blindness are predicted by the strength and stability of visual representations. *Neurosci. Conscious.* **2019**, niy010 (2019).

94. Fedorenko, E., Duncan, J. & Kanwisher, N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci.* **110**, 16616–16621 (2013).

95. Cai, W., Taghia, J. & Menon, V. A multi-demand operating system underlying diverse cognitive tasks. *Nat. Commun.* **15**, 2185 (2024).

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

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

98. 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).

99. 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).

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

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

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

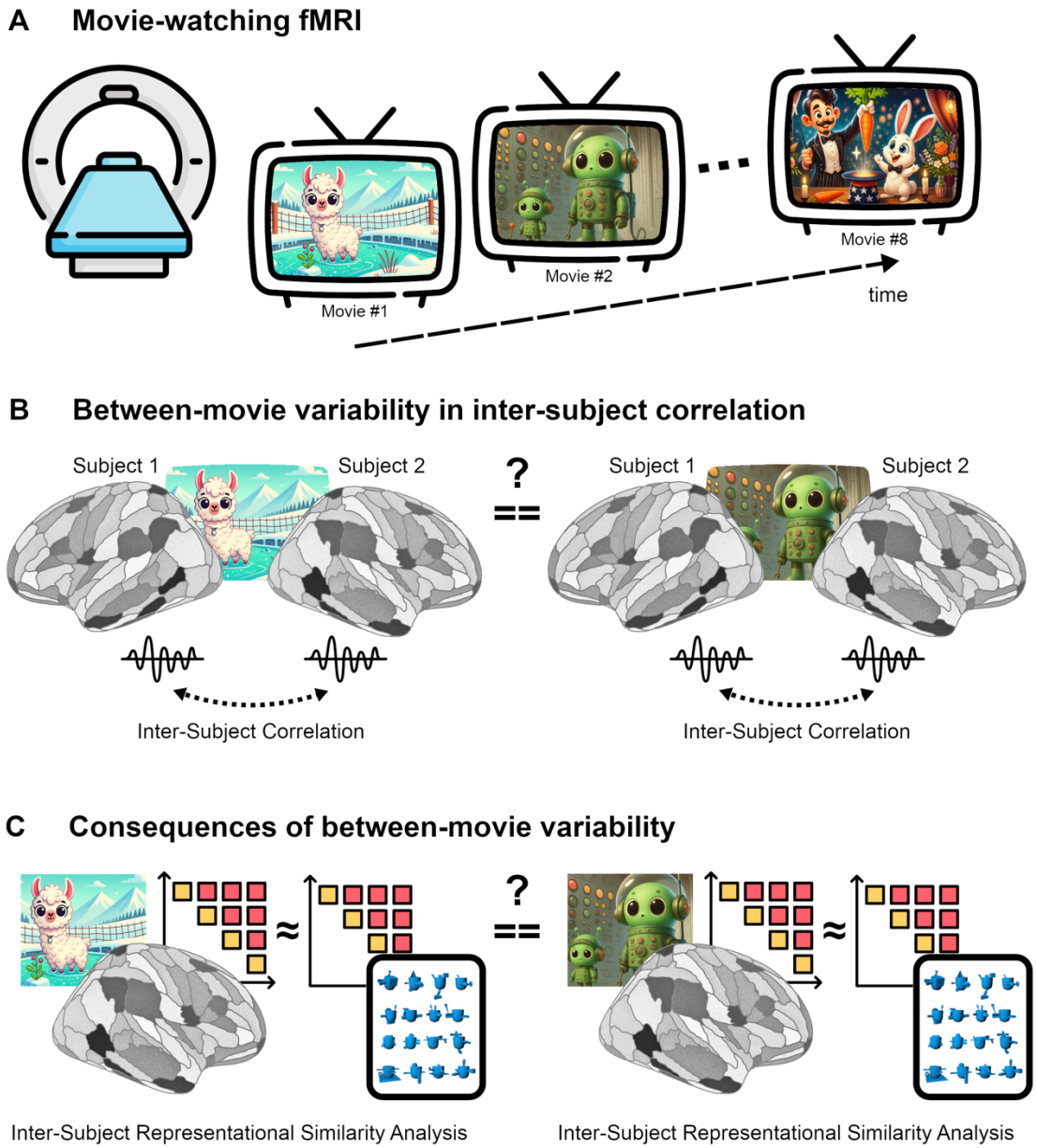


Figure 1. Overview.
A) Subjects (*N* = 112) viewed eight animated movies during fMRI scanning. The movies were selected to include diverse object categories, such as humans, plants, tools, toys, and food. **B)** We quantified the variability in inter-subject correlation (ISC) between movies across the brain. For each subject and movie, we extracted a preprocessed and spatially averaged fMRI time series from 210 cortical parcels. ISC was computed for each parcel by correlating the time series between all subject pairs. **C)** To assess the consequences of between-movie variability for generalizability, we examined whether trait- or state-like differences in ISC were consistent across movies. Specifically, we used inter-subject representational similarity (IS-RSA) to relate neural synchronization during movie watching to the similarity of subjects' conceptual-semantic representations of novel objects ("Fribbles", in blue), with analyses conducted separately for each movie.

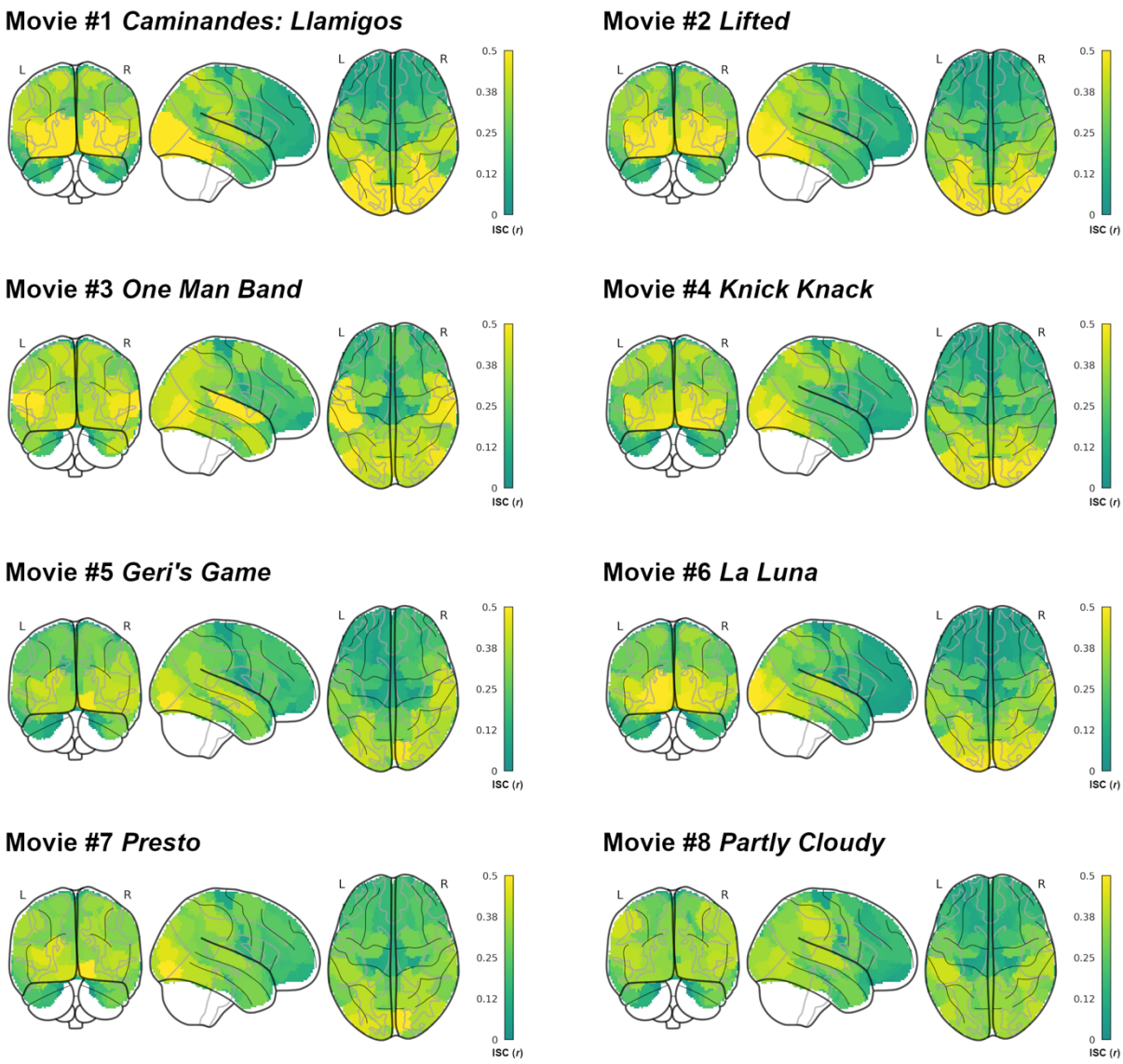


Figure 2. Inter-subject correlation for individual movies.
Spatial distribution and magnitude of inter-subject correlations (ISC) across the cortex, shown separately for each of the eight movie clips.

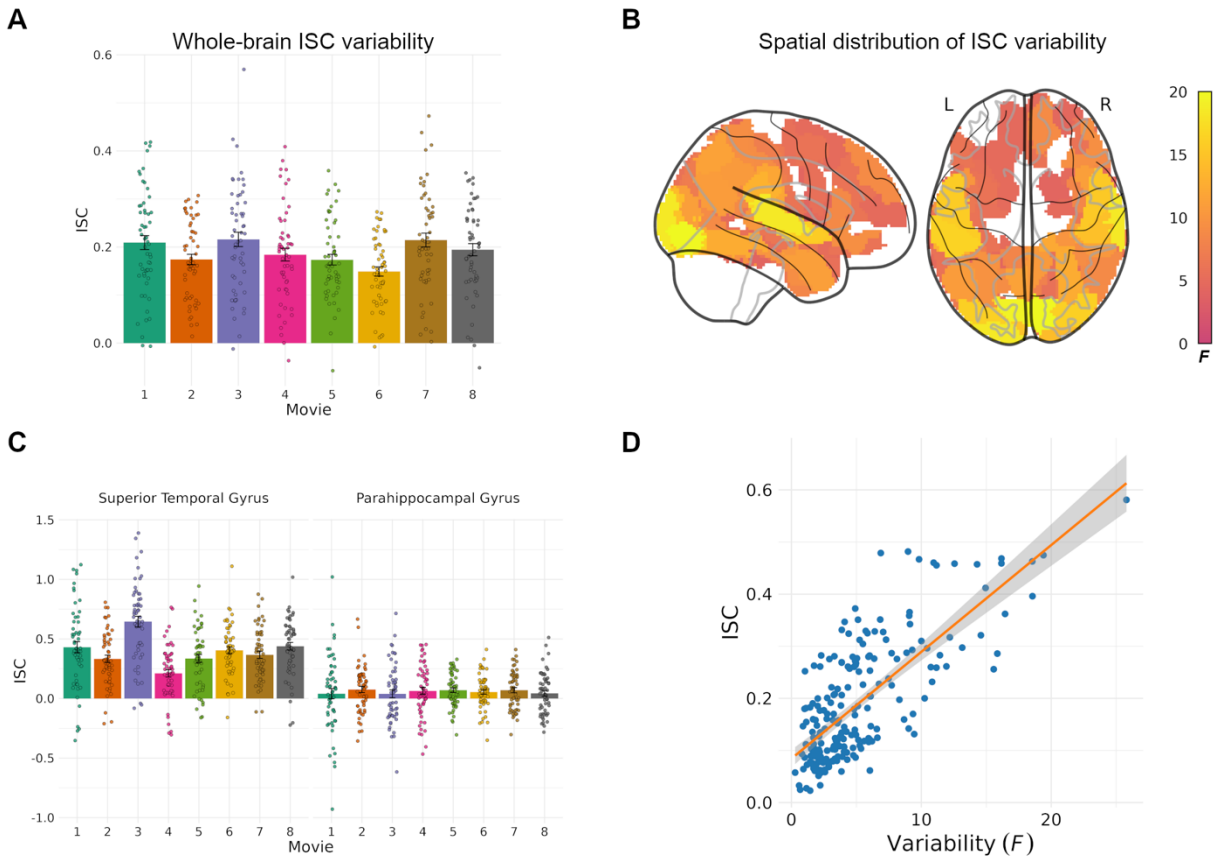


Figure 3. Variability in inter-subject correlation across movies.
A) Mean and standard error of Fisher *z*-transformed whole-brain inter-subject correlation (ISC) values for each of the eight movies. Whole-brain ISC was computed by averaging ISC values across all cortical parcels for each subject and movie. A repeated-measures ANOVA revealed a significant main effect of Movie, indicating considerable variability in whole-brain ISC across movies. **B)** Spatial distribution of between-movie ISC variability across the brain. A significant main effect of Movie was found for 97 out of 210 cortical parcels (46.2%). The highest variability was observed in primary and associative visual cortices, as well as in primary and associative auditory cortices. Significant variability was also found in higher-order regions such as the precuneus and inferior parietal lobule. **C)** Brain regions with high between-movie variability, such as the superior temporal gyrus, tended to show relatively higher ISC levels, while regions with low variability, such as the parahippocampal gyrus, showed lower ISC. **D)** A strong positive correlation between ISC level (y-axis) and between-movie variability (x-axis) across brain regions indicates that regions with higher ISC also exhibit greater variability across movies.

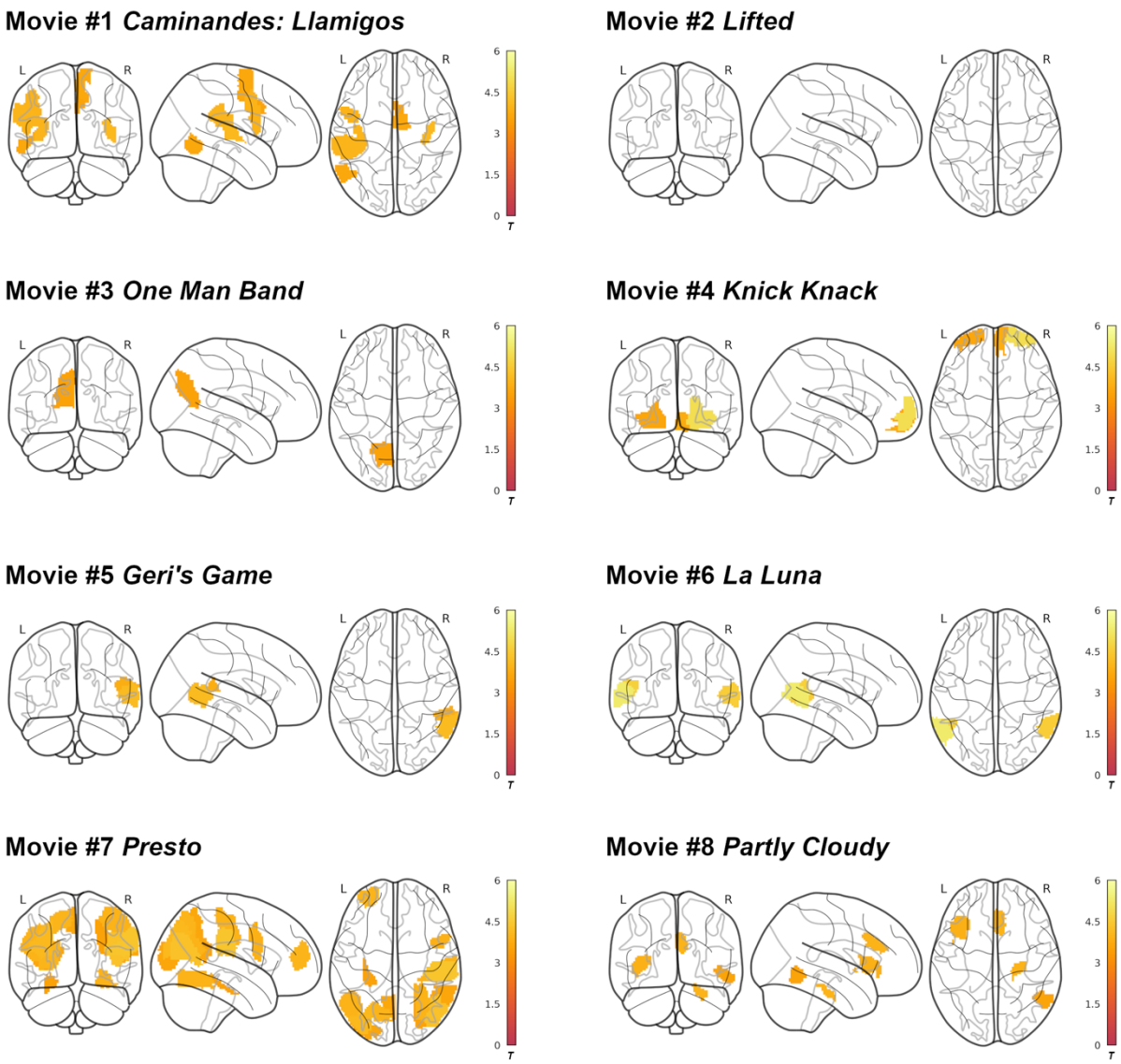
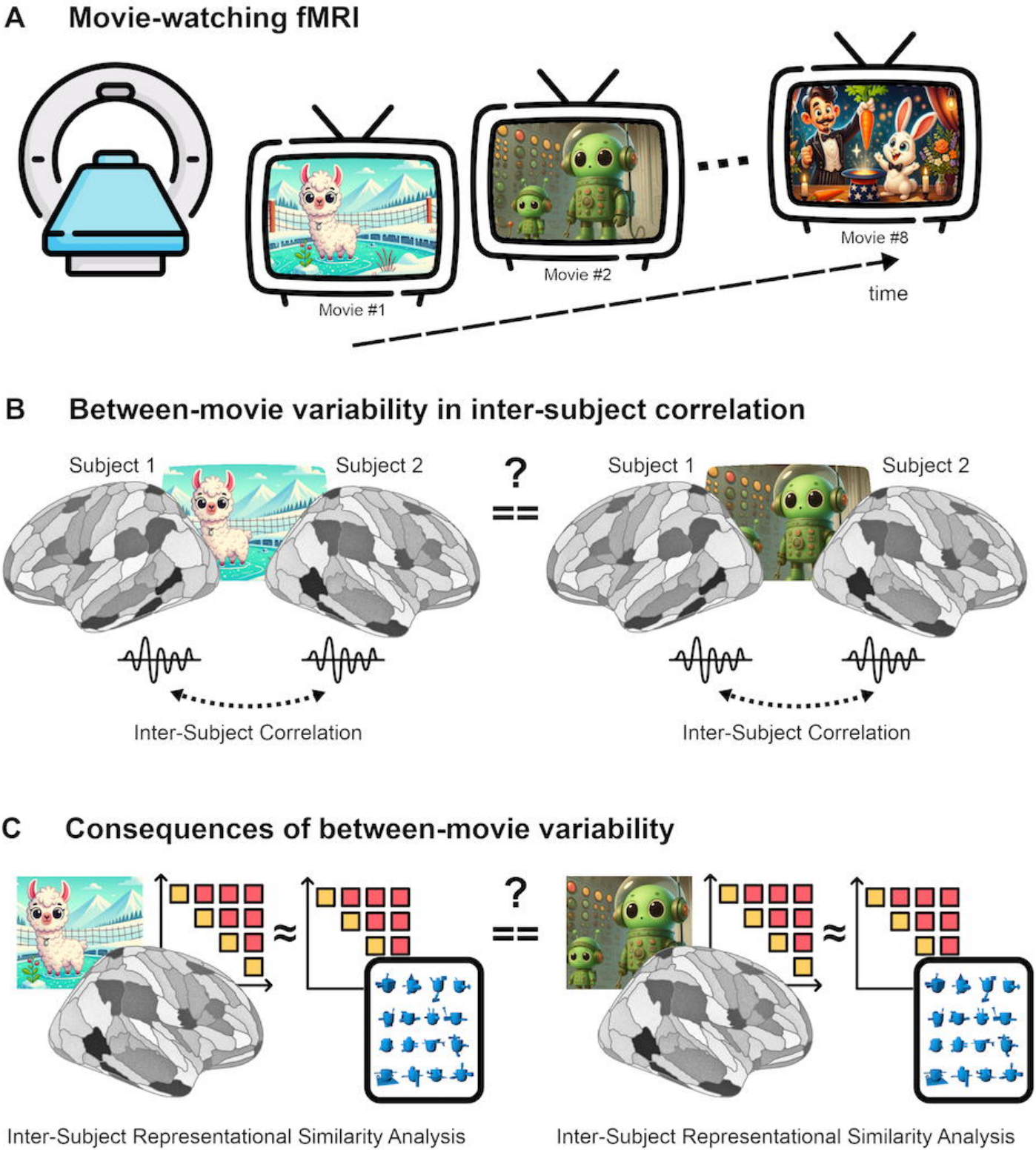
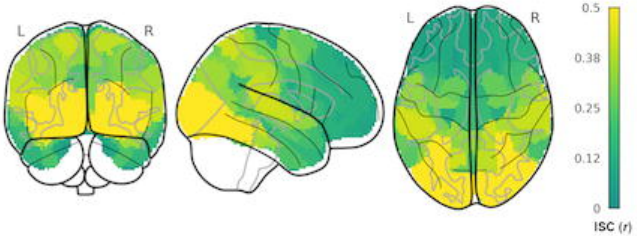


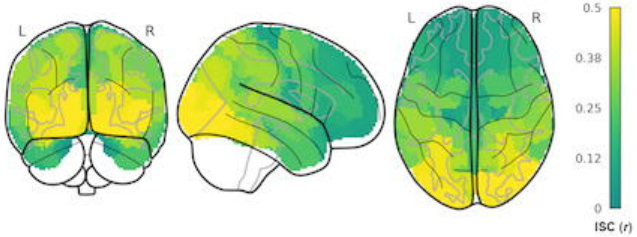
Figure 4. Between-movie ISC variability leads to varying IS-RSA associations between behavior and neural synchronization.
We used inter-subject representational similarity analysis (IS-RSA) to identify brain regions where similarity in conceptual representations of novel objects was associated with similarity in neural synchronization during movie watching. The results revealed distinct sets of brain regions for each movie (family-wise error-adjusted $p < 0.05$). These findings indicate that the association between behavior and ISC varies substantially across stimuli, limiting the generalizability of IS-RSA results based on a single movie clip.


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

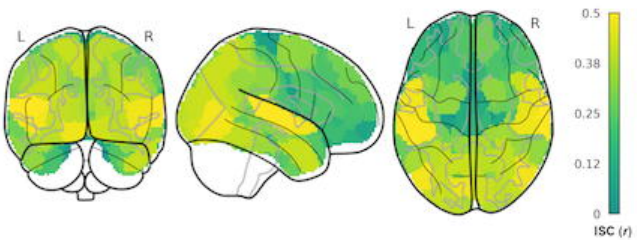
Movie #1 *Caminandes: Llamigos*



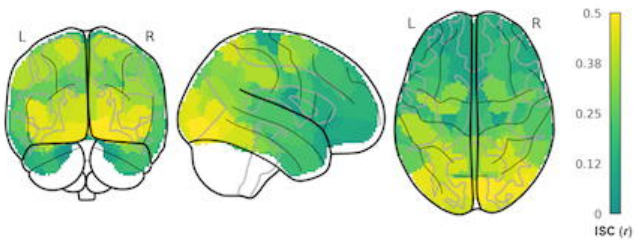
Movie #2 *Lifted*



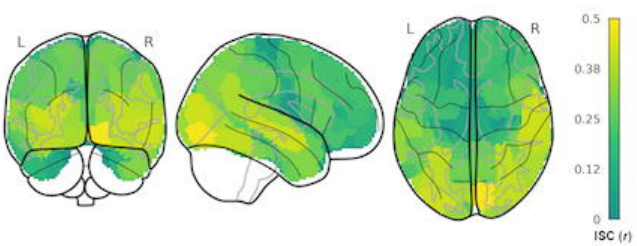
Movie #3 *One Man Band*



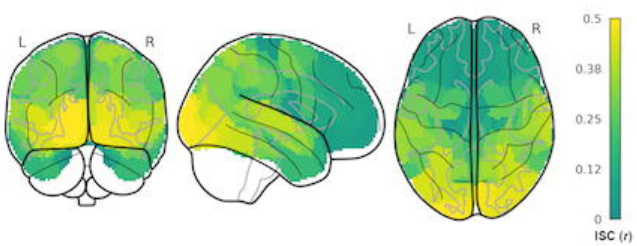
Movie #4 *Knick Knack*



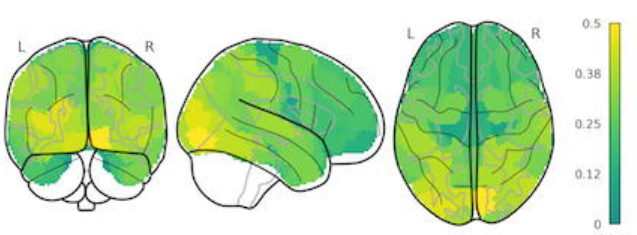
Movie #5 *Geri's Game*



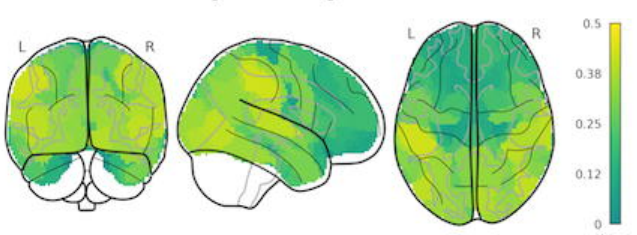
Movie #6 *La Luna*

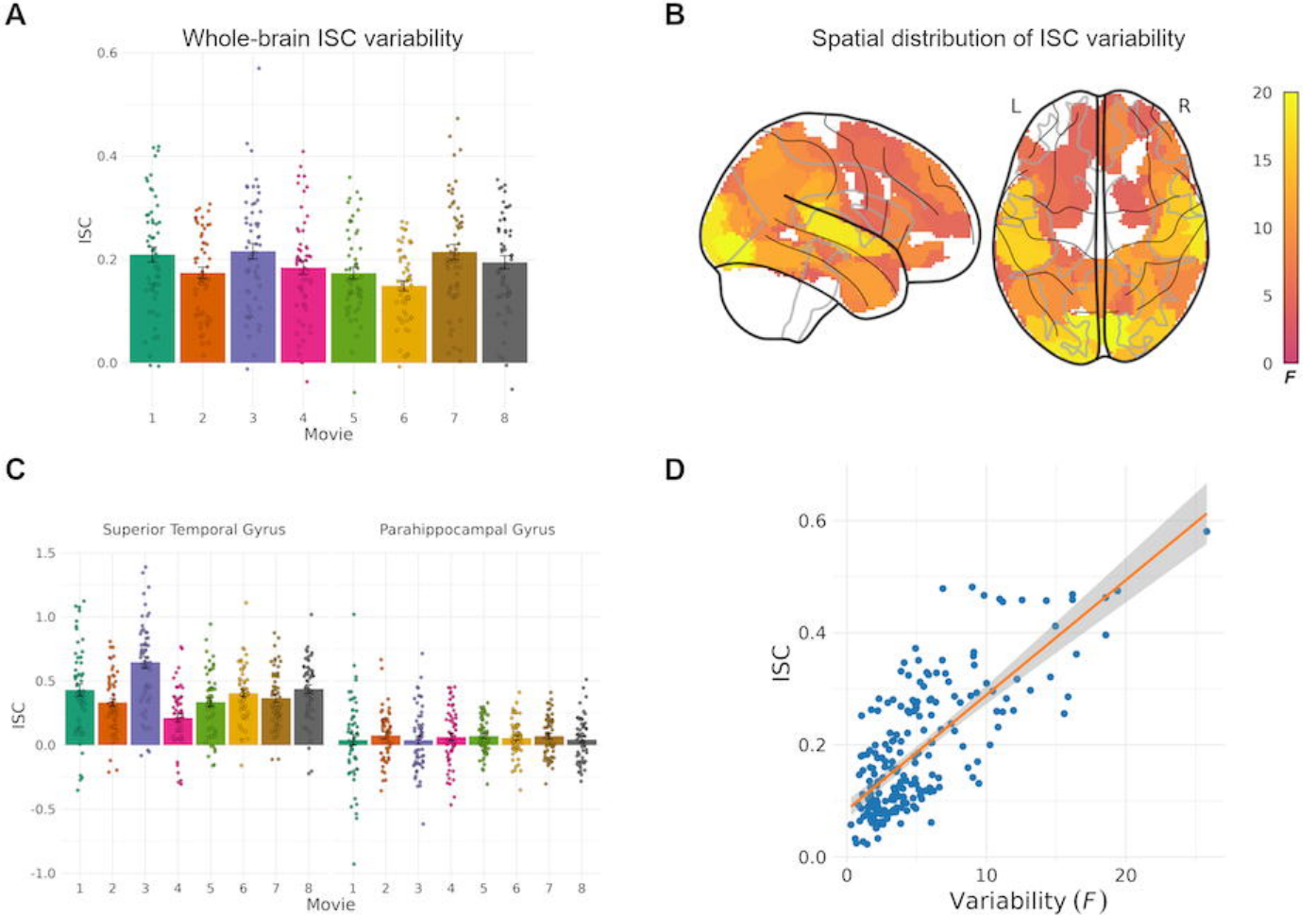


Movie #7 *Presto*

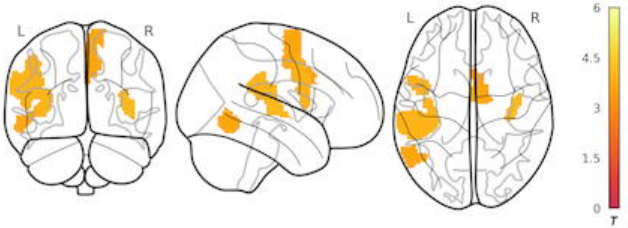


Movie #8 *Partly Cloudy*

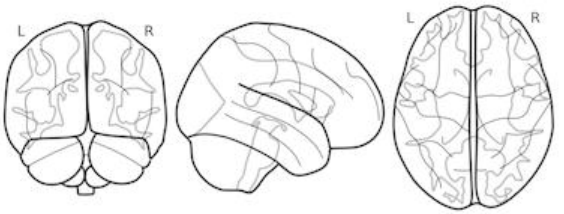



A
Whole-brain ISC variability
ISC
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
Movie
B
Spatial distribution of ISC variability
L
R
20
15
10
5
0
F
C
Superior Temporal Gyrus
Parahippocampal Gyrus
ISC
1.5
1.0
0.5
0.0
-0.5
-1.0
1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8
Movie
D
ISC
0.6
0.4
0.2
0.0
0
10
20
Variability (F)

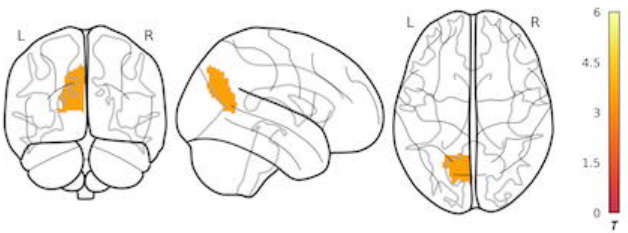
Movie #1 *Caminandes: Llamigos*



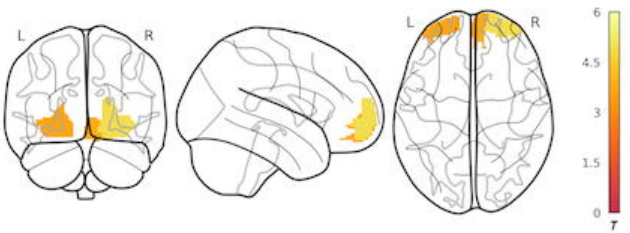
Movie #2 *Lifted*



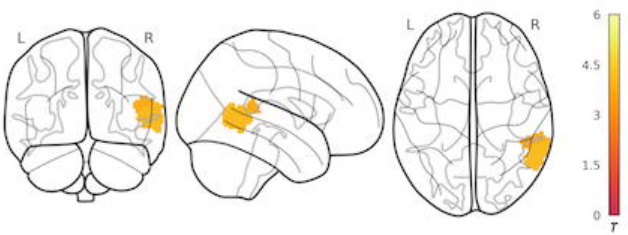
Movie #3 *One Man Band*



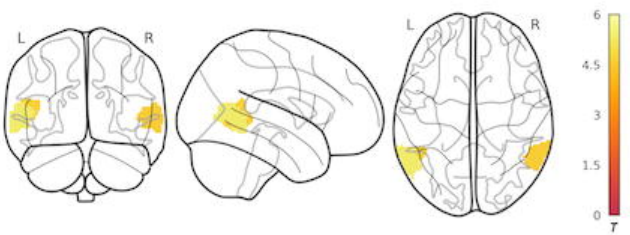
Movie #4 *Knick Knack*



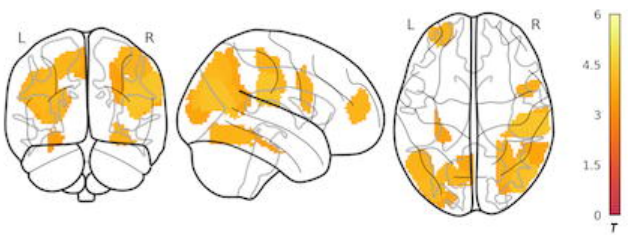
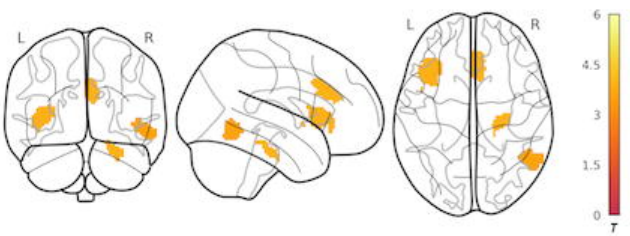
Movie #5 *Geri's Game*



Movie #6 *La Luna*



Movie #7 *Presto*



Movie #8 *Partly Cloudy*