



## UvA-DARE (Digital Academic Repository)

### Where and how our brain represents the temporal structure of observed action

Thomas, R.M.; De Sanctis, T.; Gazzola, V.; Keysers, C.

**DOI**

[10.1016/j.neuroimage.2018.08.056](https://doi.org/10.1016/j.neuroimage.2018.08.056)

**Publication date**

2018

**Document Version**

Final published version

**Published in**

NeuroImage

**License**

CC BY-NC-ND

[Link to publication](#)

**Citation for published version (APA):**

Thomas, R. M., De Sanctis, T., Gazzola, V., & Keysers, C. (2018). Where and how our brain represents the temporal structure of observed action. *NeuroImage*, 183, 677-697. <https://doi.org/10.1016/j.neuroimage.2018.08.056>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



## Where and how our brain represents the temporal structure of observed action



R.M. Thomas<sup>a,b,1</sup>, T. De Sanctis<sup>a,d,1</sup>, V. Gazzola<sup>a,c,2,\*</sup>, C. Keysers<sup>a,c,2,\*</sup>

<sup>a</sup> The Netherlands Institute for Neuroscience, An Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), Amsterdam, the Netherlands

<sup>b</sup> Department of Psychiatry, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands

<sup>c</sup> Brain & Cognition, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

<sup>d</sup> PharmAccess Foundation, Amsterdam, Netherlands

### ARTICLE INFO

#### Keywords:

Action observation  
EEG  
fMRI  
ISC  
Hebbian learning

### ABSTRACT

Reacting faster to the behaviour of others provides evolutionary advantages. Reacting to unpredictable events takes hundreds of milliseconds. Understanding where and how the brain represents what actions are likely to follow one another is, therefore, important. Everyday actions occur in predictable sequences, yet neuroscientists focus on how brains respond to unexpected, individual motor acts. Using fMRI, we show the brain encodes sequence-related information in the motor system. Using EEG, we show visual responses are faster and smaller for predictable sequences. We hope this paradigm encourages the field to shift its focus from single acts to motor sequences. It sheds light on how we adapt to the actions of others and suggests that the motor system may implement perceptual predictive coding.

### 1. Introduction

The capacity to perceive and predict actions performed by others is fundamental to proper social interactions. Over the past few decades, much research attention has been devoted to identifying the neural mechanisms that underlie the processing of simple acts such as grasping, reaching, breaking, and performing simple gestures. Electrophysiological work on non-human primates has identified that some of the neurons active while participants perform simple acts are also active when observing (or hearing) similar acts performed by others. These neurons, called ‘mirror neurons’, were originally identified in ventral premotor region F5 and in the rostral inferior parietal region PF/PFG (Gallese et al., 1996; Umiltà et al., 2001; Kohler et al., 2002; Keysers et al., 2003; Fogassi et al., 2005). Later studies have described neurons with such mirroring properties in (a) somatosensory cortices (particularly in SII and adjacent sectors of SI Hihara et al., 2015), (b) the dorsal premotor cortex (Cisek and Kalaska, 2004; Tkach et al., 2007), and (c) to a lesser extent, the primary motor cortex (Dushanova and Donoghue, 2010; Kraskov et al., 2014; Vigneswaran et al., 2013). Our current estimate of the mirror neuron system – i.e. the network of brain regions with neurons rendered active during both the observation and performance of specific actions –

comprises all these regions. Whether such mirror neurons exist elsewhere in the primate brain remains unanswered, as systematic experiments to examine the issue remain to be carried out. The firing of individual mirror neurons contains information that will permit accurate classification of the acts performed by others (C Keysers et al., 2003). This work has led to the idea that isolated observed or heard acts are processed, at least in part, by recruiting somatosensory-motor representations of the monkey's own actions (Gallese et al., 2004; Rizzolatti and Sinigaglia, 2010; Umiltà et al., 2001). A large number of neuroimaging studies in humans have identified an *action observation network* triggered by the observation of such simple acts (for ALE meta-analyses of these studies see for instance Caspers et al., 2010; Grosbras et al., 2012; Molenberghs et al., 2012). A smaller number of studies have tested the same participants during both their observation and execution of manual actions. These studies identified a network of voxels involved in both conditions (e.g. Arnstein et al., 2011; Buccino et al., 2004; Dinstein et al., 2007; Filimon et al., 2007; Gazzola and Keysers, 2009; Grèzes et al., 2003; Simos et al., 2017; Valchev et al., 2016). We shall henceforth refer to this network as the Action Observation-Execution Network (AOEN). The AOEN network includes (a) the presumed human homologue of the brain areas in which mirror neurons have been found in monkeys (vPM, dPM,

\* Corresponding authors. Netherlands Institute for Neuroscience, Meibergdreef 47, 1105BA Amsterdam, the Netherlands.

E-mail addresses: [v.gazzola@nin.knaw.nl](mailto:v.gazzola@nin.knaw.nl) (V. Gazzola), [c.keysers@nin.knaw.nl](mailto:c.keysers@nin.knaw.nl) (C. Keysers).

<sup>1</sup> Equal contributions.

<sup>2</sup> Equal contributions.

<https://doi.org/10.1016/j.neuroimage.2018.08.056>

Received 14 November 2017; Received in revised form 20 August 2018; Accepted 23 August 2018

Available online 28 August 2018

1053-8119/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SI, SII, PF/PFG) and (b) a number of regions that have not yet been systematically explored for the presence of mirror neurons in monkeys (in particular, the cerebellum, SPL, SMA and regions of the visual cortex such as V5 and EBA). Pattern classification analyses have confirmed that the pattern of brain activity in premotor, inferior parietal and somatosensory cortices does contain information that could help the organism perceive which motor act someone else performed (Etzel et al., 2008; Oosterhof et al., 2010). Disturbing activity in the somatosensory-motor nodes of this AOEN (SI, IPL, PM) leads to deficits in the processing of observed actions (for recent reviews see Avenanti et al., 2013; Keysers et al., 2018; Urgesi et al., 2014). Together, these findings suggest that humans also recruit brain regions associated with the planning, execution and somatosensation of their own actions in their perception and interpretation of the actions of others.

In contrast, we know very little about **where** and **how** the brain represents knowledge and expectations about **sequences** of acts, e.g. preparing breakfast (Grafton and Hamilton, 2007; Kilner and Frith, 2008; Thioux et al., 2008). Intelligent participation in coherent action sequences inevitably requires information that goes beyond the sum of the knowledge about the individual acts that go into their making. Representing a sequence of acts entails representing the order in which the acts were performed. Such *ordinal* information is critical to predicting actions that people are likely to perform as the follow up to a previous step. This prediction, in turn, is crucial to an intelligent agent's proactive planning of reactions to the that follow up. In this paper, we shall present the experimental evidence we have gathered about both the areas and the manner in which this knowledge is represented in the brain.

To explore **where** the brain encodes sequence level information, we localized regions responding differently to acts in a logical sequence (e.g. grasping a bun, cutting the bun, buttering the bun) and in a random sequence. Some scientists (e.g. Brass et al., 2007; Caramazza et al., 2014; Kilner and Frith, 2008) have argued that such higher-level information is more likely to be represented in the Theory of Mind (ToM) network than in the motor system. Systematic reviews of studies looking at reasoning about the mental states of others have revealed a core network composed of the medial prefrontal and rTPJ that are consistently activated whenever participants are reasoning about mental states of others irrespective of the task- and stimulus format (Mar, 2011; Schurz et al., 2014). There are some, including us, who suggest that the AOEN could represent sequence-level information. We base our suggestion on insights from experiments on monkeys showing that mirror neurons in the motor system are sensitive to expectations about upcoming actions (Fogassi et al., 2005; Umiltà et al., 2001). This is also in line with observations that premotor cortices do represent sequences of stimuli in other domains (Fiebach and Schubotz, 2006; Schubotz and von Cramon, 2001; Schubotz et al., 2004). When we act, we can see our own actions unfold in our perceptual space, so we can surmise that Hebbian learning in the synapses mutually connecting our visual and motor systems would encode the transitional probabilities across individual motor acts, and thereby enable our AOEN to represent sequence-level information and anticipation in a predictive coding framework (Keysers and Gazzola, 2014). Indeed the possibility that the AOEN is involved in such prediction is corroborated by recent experiments that show that virtual lesions to premotor cortices (Avenanti et al., 2017; Makris and Urgesi, 2015) or neurological lesions to the premotor, somatosensory or inferior parietal cortices (de Wit and Buxbaum, 2017) interferes with our ability to predict actions in a sequence.

Lerner et al. (2011) suggests a powerful experimental method to investigate this issue. They took a story and presented it to participants once in its intact form, or then after cutting it at the spaces between words and randomizing the order of the words. If brain regions are sensitive only to word-level information, randomizing the order of the words in the story should not alter brain activity. The hypothesis was that, if brain regions respond to higher, sentence- or paragraph-level information, then, randomizing the order of the words should destroy that information and reduce the efficacy of brain activity. Brain activity

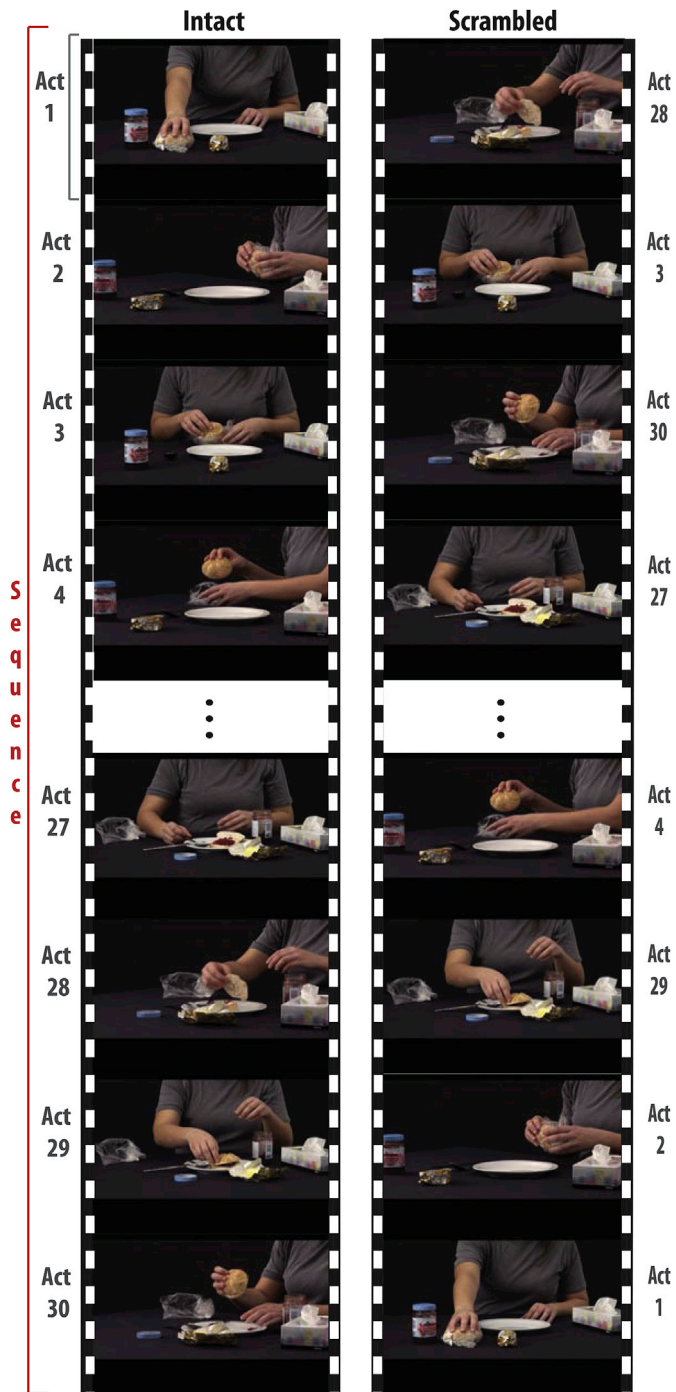
was then analysed using inter-subject correlations (ISC) (Hasson et al., 2012). ISC maps information about a stimulus in the brain in a model free fashion based on a simple logic. If a voxel has no information about a stimulus, its activity reflects spontaneous activity and will not be correlated in time with that of other participants exposed to the same stimulus. If a voxel's activity is strictly determined by a stimulus, activity across witnesses of the stimulus will be similar, and the inter-subject correlation will be significant. If so, the higher the temporal correlation between subjects with respect to a voxel, the more evidence we have of that voxel's ability to contain information about the stimulus. By comparing ISC of the intact and scrambled sentences, Lerner et al. identified brain regions that show evidence of significant additional information/correlation when sentence level information was preserved, i.e., when the sentences were presented intact, than when sentence level information was degraded, i.e., when the words were presented in a random order.

Here we adapted this approach to localize brain regions containing action sequence-level information. We recorded movies of routine actions lasting approximately 1 min (Table 1). We then measured brain activity using fMRI in 22 participants while they viewed intact movies that contain sequence- and act-level information. Then we presented the same movies disjointed at the points of transition between acts, and with the order of the acts randomized. We also measured brain activity while participants viewed these scrambled movies containing the same act-level information, but with perturbed sequence-level information (Fig. 1). We then localized brain regions that had significantly different ISC values for the intact and scrambled movies to identify regions involved in processing sequence-level information. It is important to note that not finding a region in this contrast does not mean that region has no role in encoding sequence-level information. In addition to the usual limitations regarding negative findings, this is because ISC identifies activations occurring at the same location and time across participants, and thus focuses on stimulus-locked processes (Hasson et al., 2012; Stephens et al., 2013). If different participants encode the sequence of the overall actions (e.g. making breakfast) at different points along the sequence, this would evade the ISC analysis, and a region could then be involved in encoding this form of sequence-level information without showing increased ISC. We will therefore supplement ISC analyses with analyses exploring average activity levels across the sequences to shed light on activity that is consistent in location across individuals but not in timing. We generated a simple excel sheet to illustrate the difference between ISC and a traditional block-design GLM (bGLM, see Supplementary Materials – ISC bGLM differences). The ISC detects stimulus-locked fluctuations of activity that occur at the same time for all

**Table 1**

List of sequences used as stimuli with total duration in seconds and number of motor acts.

	Action	Seconds	Acts
1	Inflating and tying a balloon.	51	27
2	Making a paper boat.	94	32
3	Preparing bread with butter and jam.	79	40
4	Sewing a button.	66	42
5	Writing a gift card.	83	39
6	Rolling a cigarette.	72	30
7	Arranging flowers in a vase.	82	39
8	Framing a picture.	112	39
9	Cleaning spectacles.	69	38
10	Cleaning a laptop screen.	46	28
11	Sending a letter.	42	34
12	Replacing battery in a torch.	51	27
13	Applying nail polish.	49	23
14	Squeezing oranges.	62	40
15	Sharpening a pencil.	83	44
16	Replacing a pillow cover.	44	35
17	Removing nail polish.	64	32
18	Preparing a sandwich.	77	27
19	Toasting bread.	65	30
20	Folding a shirt.	38	20



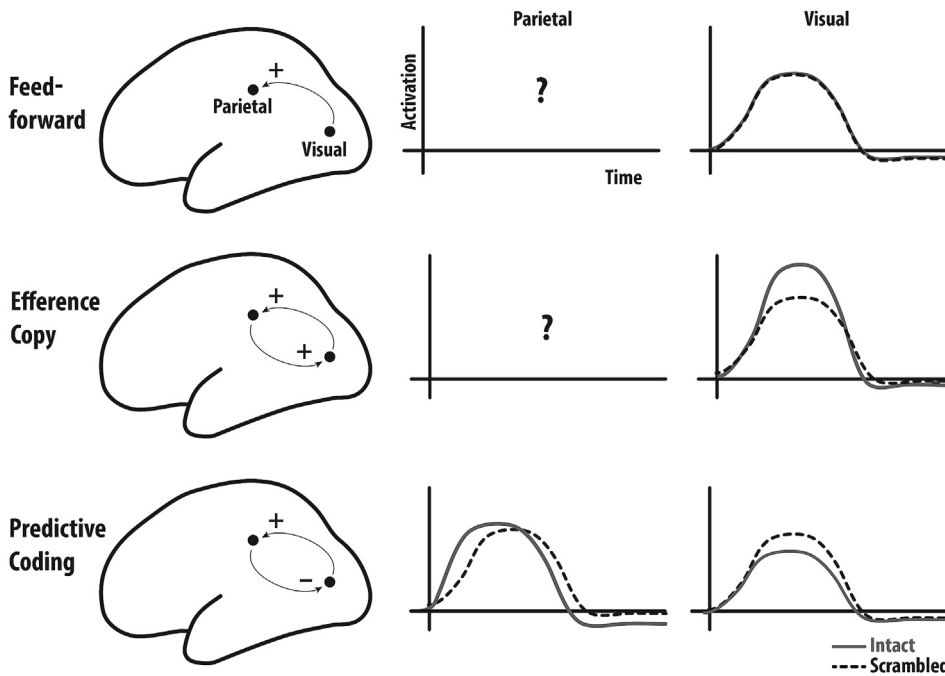
**Fig. 1. Stimulus used in the study.** A movie of a familiar action (e.g. preparing a bun for breakfast with butter and jam) is shown in an intact (left) and scrambled (right) version. Both versions contain the exact same individual acts (slicing the bun, spreading the jam, etc.), but in a different order. Note the 45° camera angle change between every two consecutive acts in both intact and scrambled sequences. This was done to ensure that the inevitable visual transients created by rearranging a sequence in the scrambled condition are also present in the intact condition, and to remove low level confounds.

participants – even if this activity does not lead to a net increase of activity. GLM, in contrast, detects net increases in activity independently of whether the timing of the increase is consistent across participants. Importantly, if a given region shows significantly more ISC for the intact version than for the scrambled version, we are justified in taking it as evidence that this region represents information about what has been

perturbed by the scrambling: the natural order of actions in the observed sequence.

Our next aim is to shed light on **how** the brain encodes sequence-level information. Anatomically, we know that the higher regions of the visual system in the temporal lobe are reciprocally connected with regions of the posterior inferior parietal lobe which in turn are connected to dorsal and ventral premotor and somatosensory brain regions (Disbrow et al., 2003; Lewis and Van Essen, 2000; Maunsell and van Essen, 1983; Nelissen et al., 2011; Pons and Kaas, 1986; Rozzi et al., 2006) (for reviews see Keysers et al., 2010; Keysers and Perrett, 2004). We can distinguish three families of models of the functional architecture of action observation based on how these models conceive of the feed-back connections back from parietal regions to the visual cortices (Fig. 2). The first family highlights the role of feed-forward connections in triggering motor programs that match visual input (Rizzolatti and Sinigaglia, 2010) without ascribing any specific function to the feed-back connections. The second family aims at explaining imitation and acknowledges the role of feed-back connections to visual regions, and assumes that these feedback connections provide excitatory efference copies that activate matching visual representations in a way akin to mental imagery (Iacoboni et al., 2001). On the basis of considerations derived from Hebbian learning and the observation that single neurons in the monkey STS are inhibited during action execution, the third family proposes that neurons in the visual cortex are inhibited by parietal predictions via inhibitory feed-back connections (Keysers and Gazzola, 2014; Keysers and Perrett, 2004) in a way akin to predictive coding models derived from a Bayesian brain perspective (Kilner et al., 2007). As inhibitory feedback cancels predictions from the visual response, the feed-forward visual information in this model becomes a representation of prediction errors rather than of what is seen in the outside world. At this point, we shall leverage the fact that these theories predict different neural activity patterns in the visual cortex as part of our strategy to shed light on the computational mechanisms involved in action observation. Purely feed-forward accounts conceive of the visual cortex only as an input stage to action perception and therefore would not be able to predict early visual areas to respond differentially to acts in their proper order and to those out of their order. Excitatory efference-copy models would suggest that the response to predicted individual acts is amplified in early visual regions by excitatory efference copies, so that early neuronal visual responses to intact sequences should be larger than response to actions in scrambled order. In contrast, inhibitory predictive coding models propose that early visual cortex essentially encodes prediction errors, and that neural activity in early visual responses should be the strongest in the case of individual acts embedded in scrambled sequences. As for the parietal node of the system, it is difficult to obtain clear predictions from the first two families of models. However, the third type of model (predictive coding) predicts that the response to acts in intact sequence should be hundreds of milliseconds faster than that to acts in scrambled sequences. This is because sensorimotor delays during re-efference are thought to wire the system so that a given action arouses expectations of the following action by priming its sensorimotor representations in the parietal cortex (Keysers and Gazzola, 2014; Keysers and Perrett, 2004). Because of its low temporal resolution fMRI is ill suited to resolve the individual motor acts embedded in our sequences or the sub-second shifts in response timing predicted by our models. Accordingly, we opted for high-density EEG to compare the evoked visual response to individual motor acts in the intact version with the responses to the scrambled version. Certain situations may complicate the predictions made by these models. For instance excitatory efference copies may down-regulate redundant sensory input, and thus minimize the expected increase of activity in early visual cortices. At the same time, inhibitory efference copies may for instance reduce neural activity for expected stimuli in pyramidal neurons (as measured with EEG) but fail to decrease the BOLD signal locally because of the metabolic costs of inhibition (but see Mangia et al., 2009).





**Fig. 2. Predictions of different action-observation models.** Feed-forward models (top) emphasize feed-forward connections from visual to parietal regions and do not ascribe a function to feed-back connections. They do not make particular predictions on the timing of parietal activations for intact and scrambled sequences (middle column) but predict that visual cortices (right-most column) respond similarly to a particular observed act embedded in the intact and the scrambled sequence. Efference-copy models (middle row) originating from imitation models suggest that feed-back connections are important and excitatory, and hence that in intact sequences, correct predictions in the parietal lobe should heighten visual responses compared to those in scrambled sequences. However, it is unclear what predictions they make regarding the timing of responses in the parietal lobe. Finally, predictive coding theories suggest that in intact sequences, the parietal lobe should show predictive responses (that thus have latencies shorter than in scrambled sequences) and inhibit responses in the visual cortex (bottom row).

## 2. Materials and methods

### 2.1. Participants

All participants were right handed as per the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision and no history of neurological or psychiatric disorder. Informed consent was provided by each participant according to the procedure approved by the ethics review board of the University of Amsterdam (2013-EXT-2847). For the fMRI experiments, 22 healthy Caucasian participants took part (11 male, 11 female, mean age  $23.3 \pm 3.46$ sd). None of the participants were excluded from the fMRI dataset. For the EEG experiment, a total of 24 participants were tested. Of these, 10 had also taken part in the fMRI experiment (5 male). The other 12 fMRI participants could unfortunately not be traced back when we decided to perform the follow up EEG experiment, and an additional 14 participants (7 male, 7 female, mean  $25.07 \pm 6.53$ sd) were recruited for the EEG experiment alone. Three of these additional EEG participants were excluded from the analyses. For two of these subjects, we found that all the channels were equally corrupted by motion artefacts and this was present in a large number of trials. This became evident by using Field-Trip's FT\_REJECTVISUAL and FT\_DATA BROWSER functions which are made available for helping with manual rejections of artefacts. The third was rejected because the impedances of the electrodes were unusually high, and the data noisy. This resulted in a final sample of 21 EEG participants (12 females, age:  $26.76 \pm 5.86$ sd).

### 2.2. Stimuli & experimental procedure

Twenty movies containing different daily actions (e.g. preparing sandwiches with butter and jam; see Table 1 for the full list) were recorded simultaneously by two video cameras (Sony MC50, 29 frames/s) at an angle of  $45^\circ$ . The videos were edited using ADOBE Premier Pro CS5 running on Windows. Each movie was subdivided into shots containing one meaningful motor act each (e.g. taking bread, opening the butter dish, scooping butter with knife, etc.). This was done on recordings from both camera angles. These motor acts (mean duration  $2\text{ s} \pm$  standard deviation  $1\text{ s}$ ) were then assembled to build two types of stimuli (Fig. 1).

For the *intact* (I) presentation, the natural temporal sequence in which the acts were recorded was maintained, but a camera angle change was introduced between every two consecutive acts by alternate sampling from the recordings of the two cameras. In the *scrambled* (S) versions, the acts remained the same, but the order of the acts was randomly rearranged, and a camera angle change was introduced between every two consecutive acts. Camera angle changes were imposed at each act transition in both types of movies to compensate for the visual transients that would otherwise be present only in the scrambled movies.

During both the fMRI and EEG experiments, participants had to watch all the 20 movies, which were presented using the Presentation software (Neurobehavioral Systems, Inc., Albany, CA, USA) in four different sessions each containing 5 intact and 5 scrambled examples, shown in a pseudo-randomized fashion, with an inter-movie interval between 8 and 12 s. No behavioural response was required during the four sessions, but participants were to carefully observe the videos. To facilitate the integration of the results in the two experiments, we adjusted the EEG setup so as to create a situation that resembles that in the fMRI setup. Specifically, the illumination of the room was dimmed down to resemble the luminance of the scanner room and the screen was placed at a distance of 120 centimetres from the participant to achieve a similar angular stimulus size.

In order to minimize the repetition effect of seeing the same movies twice, for the ten participants that took part in both the fMRI and EEG experiment, a temporal interval of 6 months was imposed between the two experiments. Besides, to ensure that participants paid attention to the movies in both the fMRI and EEG experiments, they were told that they would be required to answer three questions (that the experimenter would pick out of 22 prepared questions) to test their comprehension of the stimuli (e.g., *Did you see roses or tulips during the movie clip? What flavour was the jam? How many batteries were used in the torch?*). Comparing the number of correct responses between the fMRI and EEG experiment suggests that participants were similarly attentive: a traditional independent sample *t*-test revealed no evidence against the null hypothesis ( $t_{(44)} = -0.024$ ,  $p = 0.98$ ) and a Bayesian independent sample *t*-test as implemented in JASP (<https://jasp-stats.org>) with default settings revealed evidence for the null hypothesis of equal performance ( $BF_{10} = 0.29$ , all  $BF_{10} < 1/3$  are considered evidence for the null).

### 2.3. fMRI acquisition

Data were acquired on a 3T Philips Achieva scanner with a 32-channel head coil. Functional images were acquired with simultaneous multi-slice excitation equal to 3, (TR = 721 ms, TE = 28 ms), 39 axial slices of 3 mm with no gap and FOV of  $240 \times 240 \times 39$  mm. Images were reconstructed offline by Recon (Gyro Tools, Switzerland, <http://www.gyrottools.com>), after which the FOV was  $120 \times 78 \times 240$  mm. For each participant a T1 weighted image of  $1 \times 1 \times 1$  mm voxels was acquired. Stimuli were projected on an LCD screen and viewed through a mirror attached to the head coil.

The entire fMRI data can be found at <https://doi.org/10.5281/zenodo.1285837>.

### 2.4. fMRI inter-subject correlation analyses

Data were pre-processed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>) and custom-built MATLAB 9.8 (Mathworks Inc., Sherborn, MA) routines. The raw voxel time courses were bandpass filtered between 0.01 and 0.2 Hz, as this is known to be the optimal band to perform ISC (Kauppi et al., 2010). At this stage the BOLD time courses corresponding to all the intact and scrambled movies presented to the subject were extracted, de-meaned (voxel- and movie-wise), and concatenated in such a way that the concatenated order would remain invariant across participants irrespective of the pseudo-random order in which they saw the movies. Before concatenating to a single 4D NIFTI file, we trimmed three TRs from the beginning and the end of each movie epoch to remove the influence of non-specific BOLD transients (Hasson et al., 2004). Each subject's 4D file was then realigned to the mean image of the time course. The T1 weighted anatomical image was then co-registered to the mean functional image and segmented. All EPI images were normalized at  $2 \times 2 \times 2$  mm resolution to the template MNI brain using the forward deformation tensor derived from the segmentation of the T1 image of that subject. The normalized images were then smoothed with an  $8 \times 8 \times 8$  mm (FWHM) Gaussian filter.

Inter-subject correlations were calculated using the ISC toolbox (Kauppi et al., 2014) and in-house MATLAB routines and SPM12.

After the pre-processing step for ISC, we had two 3D time courses per subject, one for intact and the other for scrambled movies. For the subject-level ISC analysis, the time course of a given voxel in subject  $i$  was correlated with the average time course of all other subjects of that corresponding voxel. This was repeated for every voxel and with all subjects, resulting in a whole-brain map of correlation values per subject (Hasson et al., 2010; Kauppi et al., 2014).

These correlation maps were then used in a second-level random effects analysis in SPM: a one-sample  $t$ -test for  $I > 0$  ( $H_0: I \leq 0$ , i.e. in each voxel, the distribution of the ISC value across the 22 participants was compared against the null hypothesis of an  $ISC_{\text{Intact}} \leq 0$ ) and  $S > 0$  (idem for  $H_0: ISC_{\text{Scrambled}} \leq 0$ ) and a paired sample  $t$ -test comparing  $I$  and  $S$  ( $H_0: ISC_{\text{Intact}} = ISC_{\text{Scrambled}}$ ) was performed at every voxel. To determine what findings are significant, we first thresholded all contrasts at  $p < 0.001$  uncorrected with a minimum cluster size of  $k = 20$  voxels. This combination ensured that all results in the  $I$  vs  $S$  contrast survive a cluster-size family-wise error correction at  $p_{\text{fwe}} < 0.05$  (see the p-values in the results table), and most in the  $I > 0$  or  $S > 0$  contrasts. Cluster size-correction, however, does not ensure that one can interpret the location of individual voxels within the cluster. To ensure that no more than 5% of individual voxels are false positives, we determined the critical  $t$ -value that ensures a voxel-wise false discovery rate of  $q < 0.05$ . Using a voxel-wise false discovery correction for multiple comparison with  $q < 0.05$  and cluster-size threshold set at  $k > 20$  voxels,  $t$ -values greater than 2.56 and 2.58 for  $I > 0$  and  $S > 0$  respectively and 3.92 for  $I$  vs  $S$  are significant. Often, this  $t_{q < 0.05}$  is less stringent than the critical  $t$  for  $p_{\text{unc}} < 0.001$  ( $t_{p < 0.001} = 3.52$ ), and we then simply used the  $t_{p < 0.001}$  threshold. If the  $q < 0.05$  was more stringent (as in the  $I$  vs  $S$  contrast), we then used the more stringent  $q < 0.05$   $t$  value ( $t_{q < 0.05} = 3.92$ ). In short,

our threshold was always  $t = \max(t_{p < 0.001}, t_{q < 0.05})$ .

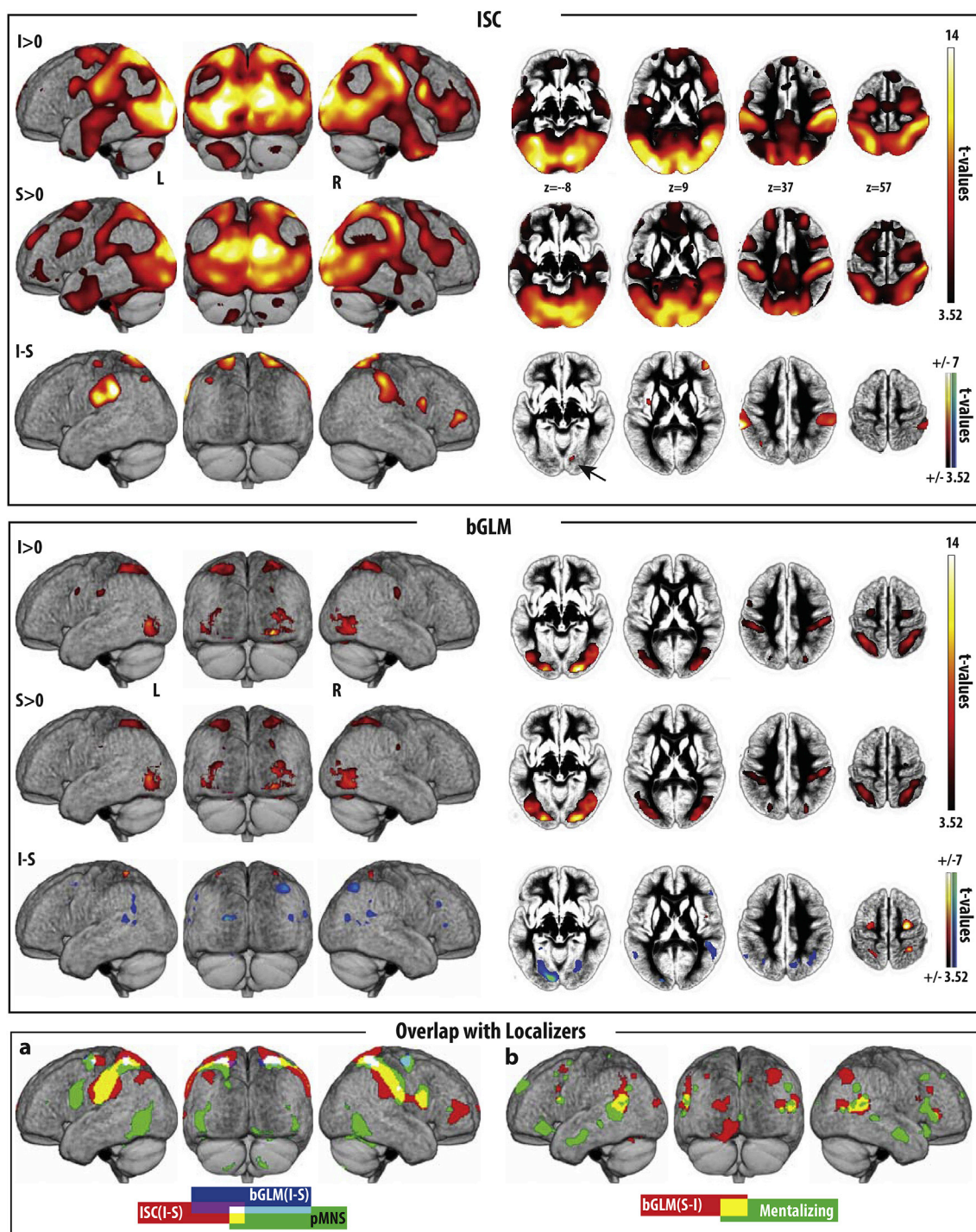
The ISC revealed that intact movies generated more synchronized activity across participants than that in the case of scrambled movies in 8 clusters (Fig. 3 ISC I–S). To determine whether these regions clustered in a smaller number of networks, we transformed these 8 clusters into regions of interest for signal extraction. Because the right parietal cluster was very large and spanned several cyto-architectonic regions at  $q < 0.05$ , we split this cluster into several sub-clusters by marginally increasing the threshold from  $t > 3.92$  to  $t > 4.1$ , which split this large cluster into three clusters (ROIs 7, 9 and 10 in Figure, 4). All other ROIs were defined at  $t > 3.92$  corresponding to  $q_{\text{FDR}; k=20} < 0.05$ , Fig. 4a and Table 2. We used Marsbar (<http://marsbar.sourceforge.net/>) to extract the time-course of each voxel within each ROI. We then calculated the Eigen-time-course in each of them. The Eigen time-course from each of these ROIs was then averaged over all participants to focus on stimulus driven activity (Simony et al., 2016) and to form a cross-correlation matrix between all the ROIs (see Supplementary Figure S1). Two clusters of activation were excluded from this ROI analysis: (1) the cerebellum, because the cluster was small (23 voxels) and located at the margin between the cerebellum and ventral visual cortex, making the interpretation as cerebellar or cortical difficult; (2) area TE3, because it was at the outermost rim of the cortex with most voxels outside of the gray-matter mask. We used the first Eigen-time-course of each of the other ROIs, despite the ROIs being relatively large, because it suffices to capture the vast majority of the variance in the signal (average = 88%, range = [78%–98%]). This correlation matrix was used as input to a *canonical multi-dimensional scaling algorithm* that positioned the ROIs on a 2D map according to the pattern of correlations between a particular ROI and others, with ROIs having similar patterns mapped closer to each other. A  $k$ -means clustering algorithm was then used to determine the number and membership of clusters in this data. To determine the number of clusters, given the relatively low number of data-points, we used the Silhouette procedure reported by Kaufman and Rousseeuw (1990). The procedure involves computing the  $k$ -means with  $k$  being set to 2, 3, 4 and 5 clusters (more clusters seemed inappropriate for 10 data-points). For each value of  $k$  we calculated the Silhouette value for each data point (a large value implies that the point is well within a cluster while low values reflect ambiguity). We repeated this procedure using the *concatenated* Eigen-time-course of all participants rather than the *averaged* Eigen-time-course. This led to the same classification in three networks for all except one ROI, viz., ROI 4, in the dorsal premotor cortex, and then switches from the blue to the red network.

This is likely to be due to the strong intrinsic connectivity between the dorsal premotor cortex and the ROIs of the red network at rest (Smith et al., 2009), which contaminates the cross-correlation matrix if averaging across participants is not performed (Simony et al., 2016).

### 2.5. fMRI general linear model analyses (GLM)

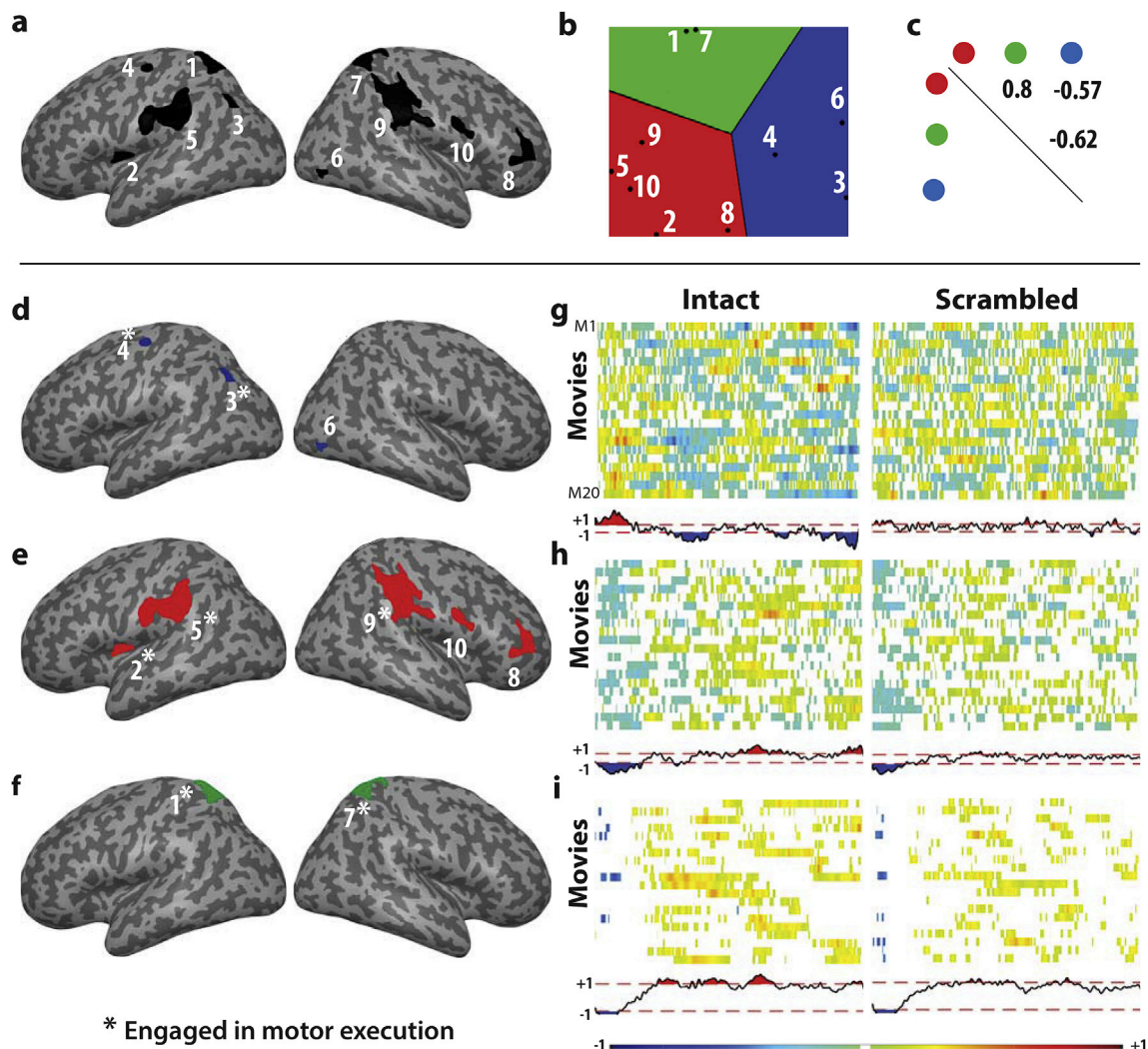
The pre-processing pipeline for the GLM analyses began with temporal filtering as in the case of the ISC. Subsequently the four different sessions were slice time corrected and realigned. The T1 image was co-registered to the mean EPI, and segmented. The EPI images were then normalized using the forward deformation tensors derived from that segmentation, written at a  $2 \times 2 \times 2$  mm resolution, and then smoothed with an  $8 \times 8 \times 8$  mm (FWHM) Gaussian filter.

In the first level analysis of the GLM, the five blocks of the intact and five of the scrambled movies were modelled as two regressors-of-interest in each of the sessions. Movement parameters estimated during realignment were included as covariates of no interest in the analysis. The regression coefficients were then used in second-level analysis in SPM: a one sample  $t$ -test for  $I > 0$  ( $H_0: I \leq 0$ ) and  $S > 0$  ( $H_0: S \leq 0$ ) and a paired sample  $t$ -test comparing  $I$  and  $S$  ( $H_0: I = S$ ) was performed at every voxel. Thresholding was performed as for the ISC results: first we imposed a  $p_{\text{unc}} < 0.001$   $k = 20$  threshold ( $t = 3.52$ ), which again ensured that most results survive a family-wise error correction for cluster size



**Fig. 3.** Regions with significant ISC (top panel) and bGLM (middle panel), and their overlap with the localizers. Each row of the top and middle panels corresponds to the contrast indicated on the left and shows lateral and posterior renders and four axial slices of the average normalized gray-matter segment of the 22 participants. Cold/warm colours represent significant negative/positive t-values. Results are shown at p-values < 0.001 and cluster-size threshold of  $k = 20$  (to impose the same  $t > 3.52$  threshold on all contrasts), but voxels not surviving a voxel wise false discovery rate (FDR) correction at  $q = 0.05$  were excluded (inclusive masking with in SPM). The bottom panel shows the results of the overlap between the ISC and bGLM analyses, and our AOEN and ToM networks. (a) Overlap across regions showing more synchrony (ISC, red) or average activation (bGLM, blue) for intact movies and the AOEN (green). No overlap was found with the ToM network, which is why this network is not shown here. (b) Overlap (yellow) between regions showing more average activation (bGLM, red) during the scrambled movies and the ToM network (green), particularly in the TPJ. No overlap was found with the AOEN, which is why this network is not shown here. See [Tables S1-S4](#), and [Tables 2 and 4-6](#) for the corresponding MNI coordinate tables. The t-maps for ISC and gGLM can also be found in .nii format in the supplementary materials together with the average anatomy of our participants.





**Fig. 4.** ISC Networks. The Eigen-time courses in the 10 regions showing ISC I-S (a) were compared using a *canonical multi-dimensional scaling algorithm* (b), in which regions with similar time courses are close to each other (see also Figure S1). A k-means clustering revealed that these regions can be summarized as 3 networks (colours in b). These networks (colored in red, green and blue separately in panels d–e) have significant correlations between them (c). (g–i) shows the average activity of these networks for each movie averaged over all participants (every row of the matrix corresponds to a movie) and the average across all movies (time course beneath each matrix). The white spaces in the matrix and the region bounded by red-dashed lines on the time course are where the ISC does not exceed that of randomized samples and is thus not significant. The ROI labels that have a star marked on top had significant co-activation with the motor-execution task (Supplementary Methods S2).

(see p-values in result tables). To interpret individual voxels, we also imposed a  $q < 0.05$  voxel-wise false discovery correction for multiple comparison. This led to a more permissive t-values of 2.81 and 2.77 respectively for  $I > 0$  and  $S > 0$ , and we thus maintained the  $p < 0.001$  threshold of  $t = 3.52$  in those cases, and  $t = 3.87$  for  $I$  vs.  $S$ , and we thus used this more stringent t-value as a threshold for that contrast.

## 2.6. EEG data acquisition

EEG data were acquired with the actiCHamp (Brain products GmbH Brain Products GmbH, Gilching, Germany) amplifier system with active electrodes. We recorded from a 128-electrode active array embedded in an elastic cap (ActiCap International, Inc.) in accordance with the 10–20 International System. In addition to the scalp electrodes, an active electrode was placed on the forehead (AFz, 25 mm above the nasion), and two electrodes on the left and right infraorbital rim to detect and clean eye movement artefacts. Impedance of all electrodes was kept below 5 k $\Omega$ . The EEG signal was digitized at a sampling rate of 500 Hz (16 bit AD converter), and a hardware high-pass filter was applied at 0.15 Hz to remove slow drifts.

Triggers were recorded at every camera change (corresponding to the transition between acts) during the movie. This was done by adding a white square on the last and first frames of each act on the side of the movie. An LED was then placed over the location of the square, and a wooden black masking frame ensured that the square was invisible to the participant. The output from the LED was then introduced as a digital channel input in the EEG recording system.

## 2.7. EEG data analysis

EEG data were analysed and pre-processed using in-house MATLAB ([www.mathworks.com](http://www.mathworks.com)) routines and the FieldTrip analysis software in MATLAB (Oostenveld et al., 2011). Eye-movement and ECG artefacts were removed using the Independent Component Analysis (ICA) procedure; a spatio-temporal ICA provides several topographical plots of ICA components, which are then manually selected for ECG and eye-movement artefacts following the criteria recommended in the FieldTrip manual ([http://www.fieldtriptoolbox.org/example/use\\_independent\\_component\\_analysis\\_ica\\_to\\_remove\\_eog\\_artifacts?s\[\]=ica](http://www.fieldtriptoolbox.org/example/use_independent_component_analysis_ica_to_remove_eog_artifacts?s[]=ica)). These components were removed and the remaining components



**Table 2**

**ISC (I–S).**

Regions with ISC Intact > Scrambled labelled using SPM Anatomy Toolbox. Results are shown, as for Fig. 3, using the most stringent threshold between the FDR correction at  $q < 0.05$  and the uncorrected at  $p < 0.001$ :  $\max(t_{FDR} = 3.93, t_{unc} = 3.53)$  and  $k = 20$ . The first column of the table also indicates whether a particular cluster survives family wise error correction at cluster level.

From left to right: the cluster size in number of voxels and FWE; the number of voxels falling in a cyto-architectonic area; the percentage of the cluster that falls in the cyto-architectonic area; the activated hemisphere (L = left; R = right); the name of the cyto-architectonic area when available or the anatomical description; the percentage of the area that is activated by the cluster; the t values of the peaks associated with the cluster followed by their MNI coordinates in mm; the number of the ROIs in Fig. 4 to which the cluster corresponds (the first cluster was split in 3 by increasing the threshold to  $t > 4.1$ ).

Cluster size	# Voxels in cyto	% Cluster	Hem	Cyto or anatomical description	% Area	Peak Information				# ROI Fig. 4
						T	x	y	z	
FWE										
ISC(I > S), $\max(t_{FDR} = 3.93, t_{unc} = 3.53) = 3.93$										
2913	416.6	14.3	R	Area PFt (IPL)	99.3					7, 9, 10
P <sub>FWE</sub> <0.000	393.2	13.5	R	Area PF (IPL)	57.9	6.18	58	-36	48	
	252.6	8.7	R	Area 2	38.6					
	204.6	7	R	Area 5L (SPL)	27.7	7.22	16	-48	78	
	183.3	6.3	R	Area 7 PC (SPL)	40.1					
	162.7	5.6	R	Area 7A (SPL)	20.8	7.15	30	-60	68	
	138.9	4.8	R	Area hIP2 (IPS)	65.5					
	125.9	4.3	R	Area 1	17.8					
	117.3	4	R	Area PFop (IPL)	51					
	117.1	4	R	Area PFm (IPL)	16.5					
	114.2	3.9	R	Area 44	18.9	5.91	62	8	26	
	87.3	3	R	Area PFcm (IPL)	26.6	5.44	60	-34	32	
	55.6	1.9	R	Area 3b	8.8	4.47	56	-10	22	
	17.4	0.6	R	Area 4a	1.6					
	14.6	0.5	R	Area hIP3 (IPS)	3.2					
	13.7	0.5	R	Area 3a	6.8					
	2.8	0.1	R	Area OP4 [PV]	0.9					
1452	448.1	30.9	L	Area PFt (IPL)	76.9					5
P <sub>FWE</sub> <0.000	257.6	17.7	L	Area PF (IPL)	49.3	8.53	-66	-36	36	
	190.5	13.1	L	Area PFop (IPL)	85.8	6.92	-68	-20	28	
	76.9	5.3	L	Area OP4 [PV]	21.3					
	68.6	4.7	L	Area 1	12.1					
	37	2.5	L	Area OP1 [SII]	9.9					
	36.4	2.5	L	Area 3b	6.5					
	16.5	1.1	L	Area PFcm (IPL)	5.1					
	5.5	0.4	L	Area 2	1					
661	342.6	51.8	L	Area 7A (SPL)	27.4	7.16	-18	-58	74	1
P <sub>FWE</sub> <0.000	193.9	29.3	L	Area 5L (SPL)	27.9					
	31.4	4.7	L	Area 7 PC (SPL)	18.4					
	21.4	3.2	L	Area 1	3.8					
306			R	Middle Frontal Gyrus		5.87	46	48	10	8
P <sub>FWE</sub> <0.000										
214			R	Callosal white matter		4.53	8	-28	26	
P <sub>FWE</sub> <0.000			L	Callosal white matter		4.48	-8	-32	22	
204			L	Rolandic Operculum		5.71	-40	0	14	2
P <sub>FWE</sub> <0.000			L	White matter		4.42	-32	-6	20	
160	69.2	42.7	L	Area PGa (IPL)	10.9					3
P <sub>FWE</sub> <0.000	13.8	8.6	L	Area hIP3 (IPS)	3	4.81	-36	-62	34	
	5	3.1	L	Area hIP1 (IPS)	1.4					
	2.8	1.7	L	Area PGp (IPL)	0.3	5.59	-38	-76	50	
96	4.9	5.1	L	Area 1	0.9	4.77	-42	-18	64	4
P <sub>FWE</sub> <0.000	1.9	2	L	Area 4a	0.2					
	1.4	1.4	L	Area 3b	0.2					
61		75.4	L	TE 3	5.2	4.65	-68	-34	4	
P <sub>FWE</sub> <0.000			L	TE 3		4.53	-68	-28	2	
50	29.4	58.8	R	Area hOc2 [V2]	2.5	4.8	12	-82	-6	6
P <sub>FWE</sub> <0.000	18.2	36.4	R	Area hOc3v [V3v]						
	2	4	R	Area hOc1 [V1]						
23	6.6	28.8	R	Lobule V (hem)	0.7	4.16	2	-56	-20	
P <sub>FWE</sub> <0.000	6.2	26.9	L	Lobule IIV	1.1					
	2.4	10.3	L	Lobule V	0.3					
	2.2	9.6	R	Lobule VII	0.4					

back-projected to the original space to obtain the “cleaned” EEG signal. Across the runs, we rejected 5 components for 11 subjects, 6 components for 5 subjects, 7 components for 3 subjects, and 8 components for 3 subjects.

Since, the task did not involve the generation of deliberate motor responses, we found almost no muscle artefacts in our data. Notch filters of 1 Hz centred on 50 Hz and its harmonics were used to remove the generic electrical cycle frequencies.

The Event-related potentials (ERPs) were calculated around each camera change marking the beginning of a new act. Each trial was defined over a window 500 ms before and 1000 ms after the camera change. We therefore had 1292 trials in total including both the scrambled and the intact movies. All trials were then averaged within condition. Because the movies are continuous, the 500 ms prior to a camera change is not a traditional baseline, but is the end of the previous act. Regarding the period after the camera change, 98% of the acts had durations longer than 650 ms. Accordingly, we focus our analyses on the first 650 ms of this epoch, which is largely unperturbed by camera changes of subsequent acts.

To determine the time points on every channel for which the ERPs from the two conditions differed significantly, we used a standard cluster-based max-sum permutation test following Maris and Oostenveld (2007) as implemented in FieldTrip while addressing multiple comparison issues (750 time points  $\times$  128 channels). For every channel and time-point the experimental conditions were compared by means of a *t*-test (with  $n$  = number of participants). All samples (channel-time pair) whose *t*-values are below a *t*-threshold corresponding to the uncorrected  $p < 0.05$  cluster-forming threshold are set to a value of zero. For all pairs exceeding this threshold, the value is first set to be the sum of its own *t*-value and those of the neighbouring channels, integrated over a time period extending from  $-25$  ms to  $+25$  ms. Then in a second step, the value is set to the maximum of the sum values across these same spatio-temporal neighbours.

Once the max-sum statistic is calculated, we need to determine its likelihood under the null hypothesis using a Monte Carlo method. Trials of the different experimental conditions (intact and scrambled) are conducted in a single set. As many trials from this combined data set as there were subjects in condition 1 are randomly drawn and placed into “pseudo-subset 1”. The remaining trials are placed in pseudo-subset 2. The test statistic (max-sum) is calculated for this random partition. The procedure for random partition and test statistics is repeated 5000 times and a histogram of the test statistic is constructed. We calculate the proportion of random partitions that result in a larger test statistic than the observed one. This proportion is the Monte Carlo significance probability, which is also our *p*-value.

If this *p*-value for a particular time-channel pair is smaller than 0.05 we conclude that the data in the two experimental conditions are significantly different at that channel and time.

For testing the significance of time-invariant parameters, the same cluster approach was used, except that the max-sum was sum applied only across neighbouring electrodes.

## 2.8. EEG source reconstruction

To capture the distributed representation of the underlying neuronal activity that resulted in the sensor-level measurements of brain activity, we performed source reconstruction using the minimum-norm estimation (MNE) method (Dale et al., 2000). MNE is an approach favoured for evoked responses and for tracking widespread activity over time. It involves solving a distributed inverse solution that discretizes the source space into locations in the brain volume using a number of current dipoles. It then estimates the amplitude of all modelled sources simultaneously to recover a source distribution whilst minimizing the overall source energy.

As part of the source reconstruction, we used an MNI template to create two geometric objects, viz., the volume conduction model and the source model. The volume conduction model determines the physics of the propagation of electrical activity through the head, which in turn depends on the conductivity of the various tissues between the source and the sensor. The source space, which will be populated with current dipoles, is the cortical sheet extracted from the anatomical image using a combination of FreeSurfer (<https://surfer.nmr.mgh.harvard.edu/>) and MNE Suite (<http://martinos.org/mne/stable/index.html>). The volume conduction and source models are then used to determine the lead fields

(from source to sensor space) using OPENMEEG (<http://openmeeeg.github.io/>). The FieldTrip EEG analysis package was used to wrap the above packages along with helper functions to construct the pipeline for source reconstruction. Noise-covariance is calculated using a time-locked analysis over the sensor space. The lead fields along with the noise-covariance was used to reconstruct the source-level activity at every time-step.

The ERPs of the two conditions (intact and scrambled) were contrasted to determine the time instances during which they differed significantly. Source reconstruction was performed at these time-points to reveal the source of the brain activity on the cortex.

## 2.9. EEG mu-suppression

To assess whether mu suppression was stronger over central electrodes, we performed a time-frequency decomposition of the EEG data from C3, C4, and Cz around each camera change. This was done for the interval  $-0.45$ – $0.85$ s relative to each camera change from frequencies ranging from 2 Hz to 40 Hz in steps of 2 Hz using FieldTrip's function FT\_FREQANALYSIS, specifically calculating the time-frequency decomposition using the MTMCOV method, which is a multi-taper time-frequency transformation based on multiplication in the frequency domain using discrete prolate spheroidal sequences (Slepian sequences) as tapers. The power at each moment and frequency was then averaged over all camera changes of the intact movies, and separately for all camera changes of the scrambled movies, to generate a single time-frequency decomposition pattern per participant and condition. Dividing the power of the intact by the scrambled decomposition yields a power ratio that should be below 1 if intact movies lead to more mu-suppression (hence less power) than scrambled movies. We then tested this ratio using a one-tailed *t*-test ( $df = 22-1$ ) separately for each time point and frequency. To correct for multiple comparisons, we used the *fdr\_bh* routine in MATLAB, with  $q = 0.05$ . To test for effects that do not vary over time, we also averaged the power over the time window, generating a single power-spectrum per participant and condition, and compared the I/S ratio against 1 (using a one-tailed *t*-test,  $df = 22-1$ ) for each frequency, using an FDR correction at  $q < 0.05$ .

## 2.10. EEG coherence analyses

To assess whether visual and parietal regions alter their connectivity in intact vs scrambled conditions, we calculated the coherence of the EEG signal between early visual cortices and the supramarginal parietal region, which the ISC analysis had revealed. We used FieldTrip to calculate the coherence between six ROIs. These ROIs are, 'V1V2V3\_Left', 'V1V2V3\_Right', 'V5\_Left', 'V5\_Right', 'LSupraMG' and 'RSupraMG'. To define the V1V2V3 ROIs, we extracted a binary image of these joined anatomical regions in MNI space using the anatomy toolbox. For V5 we used the same approach. For the supramarginal ROI ('SupraMG'), we used ROI 5 and 9 in Fig. 4E. These volume masks were mapped to the source space (nodes in the tessellation of the brain). Time courses at these source locations were reconstructed using a beamformer following the LCMV method used in FieldTrip. Sources belonging to the same mask were then averaged to obtain a mean time course per ROI and participant in the source space. These time courses were then analysed to obtain coherence spectra between the signals of pairs of ROIs. Five 200 ms time-windows were analysed in the range from  $-200$  to  $+800$  ms relative to the camera change, and coherence was calculated in 2 Hz steps from 2 to 60 Hz with a resolution of 2 Hz. Using the multi-taper FFT method implemented in FieldTrip, we obtained one spectrum per subject, per time-window, for every pair of the 3 ROIs in the left, and for every pair of the 3 ROIs in the right hemisphere. A *t*-test was then used to compare the coherence in the intact and the scrambled conditions across the 22 participants at each frequency and time window. An FDR correction was then used to correct for multiple comparisons across 5 time-windows and 30 frequencies at  $q < 0.01$ .

### 3. Results

#### 3.1. Intact movies show higher ISC

The ISC analyses revealed that both intact and scrambled movies (Fig. 3 rows one and two, respectively) induce widespread synchronization across viewers. As might be expected, in both cases the visual cortices show high ISC reflecting the stimulus-locked nature of their responses. We also see significant ISC in parietal and premotor regions. The third row depicts the contrast in neural responses to intact and scrambled (I–S) stimuli. There were no significant voxels for which the scrambled movie shows a higher ISC than the intact movies. On the other hand, a number of areas show higher ISC for intact movies. This included large clusters in the parietal lobe that extended into the L/R postcentral gyrus (including BA2, BA1 and BA3a,b), L/R superior (area 5L and 7A in particular) and inferior parietal lobule (PF/PFt in particular and a left-lateralized cluster extending along the intraparietal sulcus and PGa) and SII/PV. Other clusters were found in the dorsal mid-insula, dorsal pre-central gyrus (including BA6), the middle frontal gyrus, temporal visual area TE, occipital visual areas (including V2/V3) and cerebellar vermis and lobules V (Table 2). All of these clusters are larger than expected by chance (i.e. cluster-wise  $p_{fwe} < 0.05$ ).

#### 3.2. Overlap between ISC (I–S), AOEN and the ToM network

To investigate the degree to which I–S overlaps with regions involved in the observation and execution of individual motor acts (the so called AOEN), we used a functional localizer scan with a separate group of participants (see Supplementary Methods S1). Briefly, it includes all voxels that are activated both (a) when participants viewed goal-directed acts more than meaningless hand movements and (b) when participants executed motor acts.

The overlapping regions include large parietal clusters in both hemispheres that include somatosensory areas (BA2 and BA1 and SII in particular), inferior parietal regions (PFt/PFop in particular) and the superior parietal lobe (more specifically areas 7A and 5L). Overlapping regions also included right ventral premotor cortices (BA44), a region in the left dorsal mid insula (Table 3; Fig. 2 bottom panel). We also calculated the percentage of overlap between AOEN and ISC (I–S) to find that 19% of all AOEN voxels show more ISC during I than S, and 38% of the ISC (I–S) contrast fell within the AOEN.

To investigate the degree of overlap with the ToM network, we used the activation-likelihood estimate meta-analysis proposed by Mar (2011) that identified regions recruited by ToM based on non-story tasks. No ISC (I–S) voxels were found to overlap with these ToM areas.

#### 3.3. GLM reveals areas differentially more active during the scrambled movies

Because ISC focuses on stimulus-locked activity, to identify regions with differential stimulus-induced activity (i.e., activity triggered by the stimuli but at different times in different participants), we analysed the data using an approach in which each movie was modelled as a block in a GLM (bGLM) to capture the overall average activity during the viewing of a movie. Results of the bGLM are shown in Fig. 3 (middle panel) for  $I > 0$ ,  $S > 0$  and  $I - S$ . Note that for  $I - S$  (Table 4), we have both positive  $t$  values (i.e.  $I > S$ ) and negative  $t$ -values (i.e.  $S > I$ ). Regions with higher average activity for I than for S (positive  $t$ -values in warm colours) included bilateral dorsal precentral clusters in BA6, a right parietal cluster encompassing BA2, BA7 and BA5L and a left parietal cluster including BA7 and 5L, which were larger than expected by chance ( $p_{FWE} < 0.05$ ). We then found a smaller cluster in the left hemisphere including BA2 and 5L that did not survive cluster extent thresholding. Of these voxels, 85% fell within the regions identified by the ISC(I–S) results. Regions with higher activity for Scrambled movies (negative  $t$ -values, cold colours in Fig. 3 middle panel; Table 4) consisted of (a) large clusters that survive a

cluster-size  $p_{fwe} < 0.05$  of at least 164 voxels in the visual cortex (V1, V3, V4, Fusiform Gyrus), (b) the temporo-parietal junction (including the superior/mid temporal gyrus PGa/PGp expanding into the most caudal parts of PF and IPS), and (c) smaller clusters that fail to survive the cluster-size FWE correction in the cuneus/precuneus, cerebellum, inferior and middle frontal gyrus (incl. BA45 and BA44).

#### 3.4. Overlap between bGLM, AOEN and ToM circuit

As for the ISC, the bGLM contrast identifying preferential activation for intact sequences (I–S) revealed no overlap with the ToM network but did overlap (as the ISC I–S did) substantially (985 voxels) with the AOEN in dorsal premotor and parietal (BA7, BA2, BA5L and IPS) regions (Table 5, Fig. 2 bottom panel).

Regions showing preferential activation for scrambled sequences (bGLM S–I) overlapped in 225 voxels with the ToM network around the temporo-parietal junction (including mainly the region PGa and superior and middle temporal gyrus), and minimally (20 voxels not surviving FWE correction) with the AOEN localizer in visual brain regions (Fusiform Gyrus) (Tables 5 and 6, Fig. 2 bottom panel).

#### 3.5. Network interactions in ISC (I–S)

To further characterize the brain regions showing sequence-level information (Fig. 3 ISC (I–S) and Fig. 4a), we explored the correlation across their time courses after averaging the time courses across all participants to isolate stimulus-triggered activity. As the right parietal cluster spanned too many cytoarchitectonic regions at the FDR-corrected threshold of  $t = 3.93$  and at FWE cluster-size correction (see first cluster in Table 2), we increased the threshold for this cluster to  $t = 4.1$ , which caused it to split into three more homogeneous regions. A  $k$ -means clustering then revealed that the 10 ROIs segregate into three main networks of activity (shown as red, green and blue in Fig. 4b) based on their cross-correlation pattern. The “red” network includes bilateral supramarginal clusters (including BA2 and PF), right inferior frontal gyrus (BA45), right precentral gyrus (BA44) and left insula. The “blue” network comprises right visual cortices (including areas V1, V2 and V3) and left angular and dorsal premotor cortices. The “green” network includes clusters in the bilateral superior parietal lobule (including 7A and 5L). Correlation between the Eigen-time courses of these three networks (Fig. 4c) was positive between the red and the green networks and negative between the other pairs.

In order to have a better understanding of the kind of information encoded in these networks (Fig. 4d–f), we plotted the Eigen time course of each of the networks for each movie averaged across all subjects. By averaging the activity across all participants, one expects a flat line if activity is not synchronized across participants. The activity should peak, i.e. be positive or negative, if the stimulus contains information that triggers activity in the same direction across participants. Fig. 4g–i shows a matrix for Intact (left) and Scrambled (right) stimuli. Each row of the matrix corresponds to the brain activity triggered by a specific movie averaged across all subjects. To enable a more compact representation, the time course of activity is shown in a colour code rather than an activation line, with time along the x-axis, and colour intensity signalling the magnitude of activation. Warm colours signify positive, and cold colours negative, activations. To identify the time points at which a movie activated a network beyond what is expected by chance, we randomly permuted the labels of all the intact and scrambled movies, and averaged the activation. This procedure was repeated a thousand times. If a movie systematically triggered brain activity, one would then expect the averaged activity before permutation to have a peak that is taller than that after randomization. We therefore blanked all the parts of the matrix in which values remained within two standard deviations of the distribution of permuted data. All the parts whose values exceeded these bounds are shown in warm or cold colours.

Finally, to examine whether there are systematic trends across all



**Table 3**

**ISC (I–S) & AOEN.** Overlap between ISC (I > S) and the AOEN localizer. The ISC (I–S) contrast was inclusively masked in spm with the AOEN network described in the Supplementary Methods S1, and thresholded as in Fig. 3 and Table 2 with the max ( $t_{FDR} = 3.93$ ,  $t_{unc} = 3.53$ ). Conventions as in Table 2.

Cluster size	# Voxels in cyto	%Cluster	Hem	Cyto or anatomical description	%Area	Peak Information				
						T	x	y	z	
ISC(I > S), max ( $t_{FDR} = 3.93$ , $t_{unc} = 3.53$ ) = 3.93										
1170	391.4	33.5	R	AreaPFt (IPL)	93.8					
$p_{FWE} < 0.000$	246.8	21.1	R	Area2	38					
	98.1	8.4	R	AreaPFop (IPL)	42.9					
	77	6.6	R	Area1	11					
	68.9	5.9	R	Area7PC(SPL)	15.2					
	59.9	5.1	R	AreaPF(IPL)	8.9	6.03	56	–34	46	
			R	AreaPF(IPL)		5.91	60	–32	46	
	47.9	4.1	R	AreaHIP2(IPS)	22.7					
	39	3.3	R	Area3b	6.2	4.42	58	–12	24	
	32.4	2.8	R	Area7A (SPL)	4.2	5.64	28	–60	64	
			R	Area7A (SPL)		5.41	24	–58	66	
	20	1.7	R	Area5L (SPL)	2.7	5.19	22	–56	68	
	12.4	1.1	R	AreaHIP3(IPS)	2.7					
	4.3	0.4	R	Area3a	2.1					
	4.1	0.4	R	AreaPFm(IPL)	0.6					
	2.5	0.2	R	AreaOP4 [PV]	0.8					
721	368	51	L	AreaPFt (IPL)	63.1					
$p_{FWE} < 0.000$	145	20.1	L	AreaPFop (IPL)	65.3	6.23	–64	–20	26	
	39.1	5.4	L	AreaPF(IPL)	7.5	7.36	–64	–34	36	
	31.6	4.4	L	AreaOP1 [SII]	8.5					
	25.1	3.5	L	Area3b	4.5					
	21	2.9	L	AreaOP4 [PV]	5.8					
	17.5	2.4	L	Area1	3.1					
	5	0.7	L	Area2	0.9					
	1.9	0.3	L	AreaPFcm(IPL)	0.6					
	179	104.8	58.5	R	Area44	17.5				
	$p_{FWE} < 0.000$			R	PrecentralGyrus		5.91	62	8	26
173			L	RolandicOperculum		5.71	–40	0	14	
$p_{FWE} < 0.000$										
99	58.1	58.7	L	Area7A (SPL)	4.6	5.29	–22	–58	64	
$p_{FWE} < 0.000$	29.1	29.4	L	Area5L (SPL)	4.2					
	11.1	11.2	L	Area7PC(SPL)	6.5					

movies, we performed a grand-average across movies, and show it below the matrices together with the bounds obtained from permuted grand-average values (dotted lines in the figure). The length of the movies had been rescaled to facilitate averaging across movies. This procedure was repeated for each of the three networks. In the case of the intact movies, finding moments of significant excursion within a movie is unsurprising, given that the regions of interest were so selected as to have strong ISC for the intact movies. That the excursions remain significant *after* averaging across movies, however, indicates that the networks become synchronized consistently at specific moments of the movies. Interesting differences were noted in the timing between the three networks. The blue network, with its visual brain regions, is the first to show activation, while the other two networks are suppressed. This situation then reverses, with the activity of the green, and finally the red network becoming positive later in the movies. When processing the scrambled movies instead none of the networks shows significant deflections, except for the red network that shows negative ISC in the first part of the movie. We did not correct for multiple comparisons across time for this analysis because the analysis aimed at accurate identification of the points at which networks are most synchronized rather than just establish that they synchronize, which would be circular given the way the ROIs were selected.

### 3.6. Comparison of ISC and bGLM results

We employed ISC and bGLM as streams of analysis leading to the identification of stimulus-locked and stimulus-induced activity respectively. To better understand the relation between these methods, we also extracted the parameter estimates of the bGLM from regions showing

increased ISC, and ISC values from regions showing altered bGLM (Fig. 5, top row). In each case, we plotted the parameter estimates (beta values) for the ISC on the x-axis and the bGLM values on the y-axis. For each ROI, we then show the value for the intact movie as a circle, and for the scrambled movie as a cross. For ROIs selected on the basis of increases in ISC, there is no point in statistically comparing the ISC value for intact and scrambled movies, as this would be circular. However, we tested for significant changes in the bGLM. The middle row of Fig. 5 shows significant changes with a solid line and insignificant changes with a dashed line. This analysis makes it clear that parietal regions of the red network (ROI 4 and 10) displayed increase in average brain activity and higher ISC in the intact than in the scrambled movies. At the same time, regions in the prefrontal cortex and visual regions (ROI 1, 6 and 8) showed reduced average activity and increased ISC. The other ROIs did not change their average activity level at all, despite changes in ISC. For ROIs selected on the basis of changes in bGLM, we tested the significance of the change in ISC. We see that most bGLM ROIs that show higher activity in the Intact condition (Fig. 4, middle column) also show higher ISC. In contrast, none of the bGLM ROIs showing more activity in the scrambled condition showed significant changes in the ISC (Fig. 4, rightmost column). This suggests that activity triggered specifically by the intact movies, as identified using a GLM, was stimulus-locked, and consequently led to increased ISC. Activity triggered specifically by the scrambled movies was instead stimulus-induced rather than stimulus-locked. The bottom row of Fig. 5 depicts the regions of overlap between ISC (I–S) and bGLM (I–S), and between ISC (S–I) and bGLM(S–I). It also shows that increases in ISC(I–S) can overlap with both bGLM increases (I–S) and decreases (S–I).

**Table 4**

**BGLM.** Regions with bGLM Intact > Scrambled, and Scrambled > Intact. Contrasts were thresholded with  $t = 3.87$  and  $t = 3.53$  respectively, corresponding to the max ( $t_{FDR} = 3.87$ ,  $t_{unc} = 3.53$ ) and max ( $t_{FDR} = 3.19$ ,  $t_{unc} = 3.53$ ). Conventions as in Table 2.

Cluster size FWE	# Voxels in cyto	% Cluster	Hem	Cyto or anatomical description	% Area	Peak Information			
						T	x	y	z
<b>bGLM(I &gt; S), max (<math>t_{FDR} = 3.87</math>, <math>t_{unc} = 3.53</math>) = 3.87</b>									
436			R	Precentral Gyrus		7.81	26	-14	58
$P_{FWE} < 0.000$			R	Precentral Gyrus		5.90	26	-10	48
260	108.9	41.9	R	Area7PC(SPL)	15	6.33	30	-46	58
$P_{FWE} < 0.002$	97.3	37.4	R	Area2	4.7				
	34.5	13.3	R	Area5L (SPL)	0.9	5.11	18	-56	60
	6.9	2.6	R	Area7A (SPL)	0.8				
	3.8	1.4	R	AreaIP3(IPS)	0.1				
236			L	Precentral Gyrus		6.37	-26	-16	52
$P_{FWE} < 0.001$			L	Precentral Gyrus		5.28	-26	-14	60
			L	Precentral Gyrus		4.92	-18	-18	60
			L	Superior Frontal Gyrus		4.47	-20	-6	64
163	88.6	54.4	L	Area5L (SPL)	12.8	5.60	-22	-50	60
$P_{FWE} < 0.010$	27.8	17	L	Area7A (SPL)	2.2				
	26.8	16.4	L	Area7PC(SPL)	15.7	6.05	-32	-48	66
50	27.9	55.8	L	Area2	5.3	5.13	-36	-38	48
$P_{FWE} > 0.24$	1.3	2.5	L	Area5L (SPL)	0.2				
<b>bGLM(S &gt; I), max(<math>t_{FDR} = 3.19</math>, <math>t_{unc} = 3.53</math>) = 3.53</b>									
1090	277.8	25.5	L	AreaOc4v [V4(v)]	38.2	6.00	-24	-76	-16
$P_{FWE} < 0.000$	265	24.3	L	AreaOc3v [V3v]	28.6	8.79	-10	-88	-10
	158.8	14.6	L	AreaFG1	62.3	4.78	-30	-68	-18
	112.9	10.4	L	AreaOc1 [V1]	5.6				
	90.6	8.3	L	LobuleVI(Hem)	4.8	4.36	-34	-64	-22
	67.6	6.2	L	LobuleVIIacrusI(Hem)	2.2				
	47.3	4.3	L	AreaOc2 [V2]	5				
	18.4	1.7	L	AreaFG4	3.1				
	5.5	0.5	L	AreaFG2	1.1				
	1	0.1	L	LobuleVI(Verm)	0.5				
			L	FusiformGyrus		4.42	-30	-62	-4
628	95.9	15.3	R	AreaPGa(IPL)	12.9	5.12	64	-48	20
$P_{FWE} < 0.000$			R	AreaPGa(IPL)		4.06	54	-50	24
	51.1	8.1	R	AreaOc4a	5.8	3.78	54	-68	12
			R	AreaOc4a		3.70	52	-70	10
	29.6	4.7	R	AreaPGp (IPL)	3				
	15.9	2.5	R	AreaPFm(IPL)	2.3				
	8.4	1.3	R	AreaPF(IPL)	1.2				
			R	SuperiorTemporalGyrus		5.89	46	-38	10
			R	MiddleTemporalGyrus		5.85	56	-50	6
			R	MiddleTemporalGyrus		5.00	44	-72	18
			R	MiddleTemporalGyrus		4.24	56	-60	8
431	58.1	13.5	R	AreaIP1(IPS)	20.1				
$P_{FWE} < 0.000$	25.9	6	R	AreaIP3(IPS)	5.7				
	24.4	5.7	R	AreaPGa(IPL)	3.3				
	15.9	3.7	R	AreaPGp (IPL)	1.6				
	5.4	1.2	R	Area7A (SPL)	0.7				
			R	Angular Gyrus		5.68	42	-66	50
			R	Angular Gyrus		5.15	36	-58	42
420	145.1	34.6	L	AreaPGa(IPL)	22.7	4.70	-54	-56	26
$P_{FWE} < 0.001$	40.5	9.6	L	AreaPFm(IPL)	7	4.55	-56	-58	40
			L	AreaPFm(IPL)		4.20	-56	-52	46
			L	AreaPFm(IPL)		3.67	-52	-60	44
	7.8	1.8	L	AreaPFcm(IPL)	2.4				
	2.4	0.6	L	AreaPGp (IPL)	0.3				
			L	SupraMarginalGyrus		5.20	-50	-50	24
			L	SuperiorTemporalGyrus		4.97	-64	-48	12
			L	SuperiorTemporalGyrus		4.44	-58	-48	18
			L	MiddleTemporalGyrus		4.64	-56	-58	10
			L	Angular Gyrus		3.95	-40	-52	26
			L	Angular Gyrus		4.03	-38	-54	28
174	96.8	55.6	R	AreaOc4v [V4(v)]	15.6	4.26	24	-76	-12
$P_{FWE} < 0.040$	39.9	22.9	R	AreaFG1	16	4.70	26	-64	-8
	32.4	18.6	R	AreaOc3v [V3v]	3.8				
164	71.9	43.8	L	AreaOc4d [V3A]	12.6	7.53	-20	-92	14
$P_{FWE} < 0.048$	34.9	21.3	L	AreaOc3d [V3d]	3.5	3.96	-10	-94	22
	12.5	7.6	L	AreaOc4lp	1.5				
	5.5	3.4	L	AreaOc1 [V1]	0.3				

(continued on next page)

**Table 4** (continued)

Cluster size FWE	# Voxels in cyto	% Cluster	Hem	Cyto or anatomical description	% Area	Peak Information			
						T	x	y	z
106			R	Cuneus		4.76	12	-70	38
$p_{FWE} > 0.166$			R	Cuneus		4.05	20	-62	36
			R	Precuneus		3.88	18	-60	34
	99	30.6	30.9	R	AreaFG4	6.3			
$p_{FWE} > 0.194$	1.3	1.3	R	LobuleVI(Hem)	0.1				
			R	FusiformGyrus		5.69	26	-44	-14
	81	27.3	33.6	R	Area45	2.6	4.40	52	32
$p_{FWE} > 0.289$			R	IFG (p.Triangularis)		4.73	42	36	0
	70	35.6	50.9	L	AreaIP3(IPS)	7.8	4.02	-30	-62
$p_{FWE} > 0.368$	9.6	13.8	L	AreaPGa(IPL)	1.5	3.78	-34	-68	46
	2.8	3.9	L	Area7A (SPL)	0.2				
	64	2.1	3.3	R	Area44	0.4			
$p_{FWE} > 0.419$			R	IFG (p.Triangularis)		4.58	44	12	22
	49		R	MiddleFrontalGyrus		4.90	50	26	32
$p_{FWE} > 0.573$			R	MiddleFrontalGyrus					
	41		L	MiddleFrontalGyrus		4.57	-42	6	52
	$p_{FWE} > 0.667$			L	MiddleFrontalGyrus		4.12	-40	10
			L	MiddleFrontalGyrus		3.85	-34	10	56
37	1	2.7	L	Area44	0.1				
$p_{FWE} > 0.716$			L	IFG (p.Triangularis)		4.11	-38	18	18
			L	IFG (p.Opercularis)		3.81	-42	10	20
	36		L	IFG (p.Opercularis)		4.38	-40	14	32
$p_{FWE} > 0.728$									

**Table 5**

**BGLM & AOEN.** Overlap between bGLM ( $I > S$ ) and the AOEN localizer. The bGLM ( $I-S$ ) and bGLM ( $S-I$ ) contrasts were inclusively masked in spm with the AOEN network described in the Supplementary Methods S1, and thresholded with  $t > 3.87$  for  $I-S$  (max ( $t_{FDR} = 3.87$ ,  $t_{unc} = 3.53$ )) and  $t > 3.53$  for  $S-I$  (max ( $t_{FDR} = 3.87$ ,  $t_{unc} = 3.53$ )). Conventions as in Table 2.

Cluster size FWE	# Voxels in cyto	% Cluster	Hem	Cyto or anatomical description	% Area	Peak Information				
						T	x	y	z	
<b>bGLM (<math>I &gt; S</math>) &amp; AOEN, max (<math>t_{FDR} = 3.87</math>, <math>t_{unc} = 3.53</math>) = 3.87</b>										
379			R	Precentral Gyrus		7.81	26	-14	58	
$p_{FWE} < 0.000$			R	Precentral Gyrus		5.9	26	-10	48	
	233	106.9	45.9	R	Area7PC(SPL)	23.5	6.33	30	-46	58
$p_{FWE} < 0.002$	95	40.8	R	Area2	14.6					
	24	10.3	R	Area5L (SPL)	3.3	5.11	18	-56	60	
	3.8	1.6	R	AreaIP3(IPS)	0.8					
	2.3	1	R	Area7A (SPL)	0.3					
204			L	Precentral Gyrus		6.37	-26	-16	52	
$p_{FWE} < 0.004$			L	Precentral Gyrus		5.28	-26	-14	60	
			L	Precentral Gyrus		4.72	-20	-18	62	
			L	Superior Frontal Gyrus		4.47	-20	-6	64	
	119	74	62.2	L	Area5L (SPL)	10.7	5.6	-22	-50	60
$p_{FWE} < 0.031$	20.9	17.5	L	Area7PC(SPL)	12.2	5.8	-30	-50	64	
	19.6	16.5	L	Area7A (SPL)	1.6					
	50	27.9	55.8	L	Area2	5.3	5.13	-36	-38	48
$p_{FWE} > 0.249$	1.3	2.5	L	Area5L (SPL)	0.2					
	<b>bGLM (<math>S &gt; I</math>) &amp; AOEN, max (<math>t_{FDR} = 3.19</math>, <math>t_{unc} = 3.53</math>) = 3.53</b>									
20	16.5	82.5	L	AreaFG1	6.5	4.58	-30	-66	-18	
$p_{FWE} > 0.908$	2	10	L	LobuleVI(Hem)	0.1					
	1.5	7.5	L	AreaFG4	0.3					

**3.7. Intact movies lead to faster and smaller EEG responses compared to scrambled movies**

Different models of the mirror neuron system inspire different predictions about responses to individual motor acts depending on whether they are embedded in intact or scrambled sequences (Fig. 2). To test these predictions, we complemented the fMRI analyses with an analysis of electrophysiological responses triggered by each individual motor act, by aligning the EEG responses to the camera changes. The EEG data have higher temporal resolution. They also overcome the limitation of fMRI,

viz., that inhibition might lead to an increase in BOLD due to the metabolic cost of the inhibition (Mangia et al., 2009). Fig. 6 shows the ERP that has been averaged across five distinct sets of EEG channels progressing from posterior to anterior channels. We found that the ERP for intact movies started rising faster than for the scrambled movies, which was visible in the difference curves (Fig. 6, insets) as an initial red phase. We found that the ERP for intact movies had lower amplitude than that for the scrambled movies later in the ERP, which is visible as a blue phase in the difference curve.

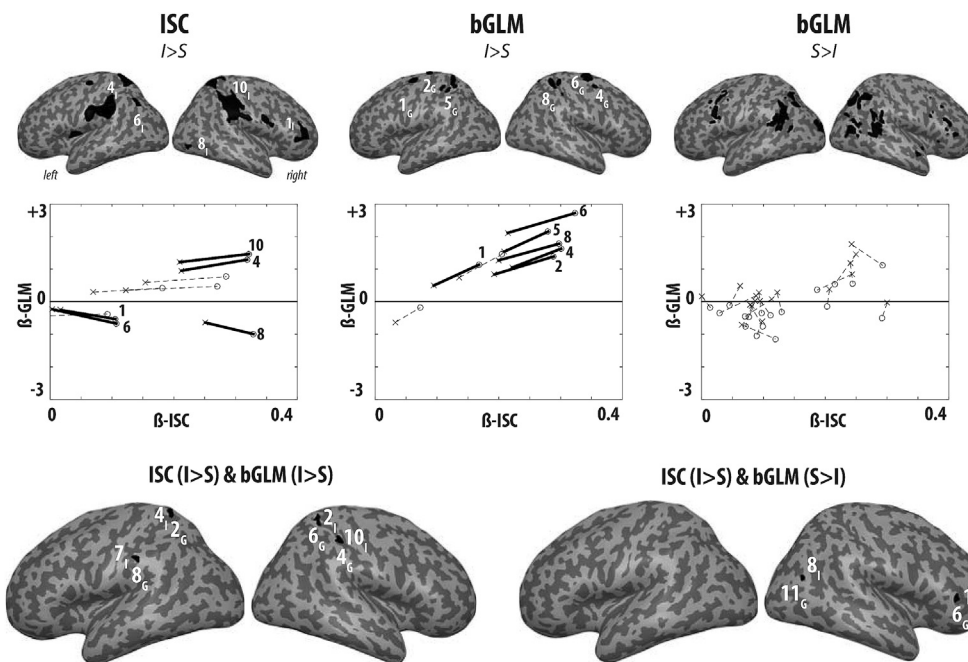
To quantify these observations, we performed two analyses. First, using



**Table 6**

**BGLM (S–I) & ToM areas.** Overlap between bGLM (S > I) and the ToM regions. The bGLM (S–I) contrasts was inclusively masked in spm with the ToM network described in Mar et al. (2011), and thresholded with  $t > 3.53$ , which corresponds to the max ( $t_{FDR} = 3.19, t_{unc} = 3.53$ ). Conventions as in Table 2.

Cluster size	# Voxels in cyto	% Cluster	Hem	Cyto or anatomical description	% Area	Peak Information			
						T	x	y	z
bGLM (S > I) & ToM, max ( $t_{FDR} = 3.19, t_{unc} = 3.53$ ) = 3.53									
225	32.6	14.5	R	Area PGa (IPL)	4.4	3.95	54	–50	22
$P_{FWE} < 0.015$	7.5	3.3	R	Area PGa (IPL)	0.8	3.85	60	–52	20
			R	Area PGp (IPL)					
			R	Superior Temporal Gyrus		5.48	48	–42	12
			R	Middle Temporal Gyrus		5.16	54	–50	8
			R	Middle Temporal Gyrus		4.24	56	–60	8
143	47.4	33.1	L	Area PGa (IPL)	0.6	5.1	–52	–52	22
$P_{FWE} < 0.075$	3.5	2.4	L	Area PGa (IPL)	0.2	4.7	–54	–56	26
			L	Area PFm (IPL)					
			L	Area PGp (IPL)					
			L	Middle Temporal Gyrus		4.62	–54	–56	10
			L	Superior Temporal Gyrus		4.44	–58	–48	18

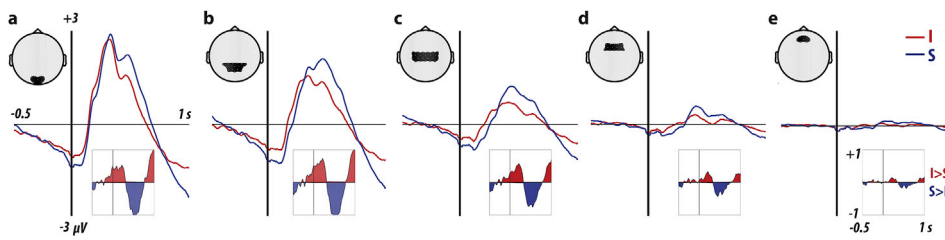


**Fig. 5. The ISC – bGLM relation.** The top row depicts the ROIs corresponding to (from left to right) ISC, bGLM (I–S) and bGLM (S–I). The middle row the parameter estimates ( $\beta$  weights) from the ISC (x-axis) and bGLM (y-axis) as a function of the Intact/Scrambled state of the movie for the ROIs shown above. Significant differences between intact and scrambled in the dimension that was not used to define the ROI is shown by solid lines, non-significant differences, by dotted lines. The bottom row shows the overlap between ISC(I–S) with bGLM (I–S) and ISC(I–S) with bGLM (S–I). The numbers correspond to the ROI and the subscript I/G indicates membership to ISC/bGLM, so that  $4_{I2G}$  indicates an overlap between ROI 4 of the ISC and ROI 2 of the GLM.

two simple models fitted separately to each participant's ERPs, we explored the presence of a shift in latency and a difference in response magnitude. In Model 1, in which Scrambled( $t$ ) = Intact ( $t-\lambda$ ), we optimized  $\lambda$  so as to reduce the residual error over the period from 0 to 300 ms in order to focus on the rising of the ERP and quantify the shift in latency. In Model 2, where Intact( $t$ ) =  $\alpha$ \*Scrambled( $t$ ), we optimized  $\alpha$  so as to reduce residual errors for  $t$  from 0 to 650 ms in order to take the entire response into account and quantify the relative response magnitude. We then compared the fitted parameters against the null hypotheses  $\lambda = 0$  and  $\alpha = 1$  across participants using a cluster-based statistic (cf. the Method Section). Fig. 7a shows a significant shift in response latency over occipital and parietal electrodes, with the intact condition leading to earlier responses (left panel) and a significant scaling over occipital, parietal and (left) frontal electrodes with the scrambled condition leading to larger responses. Second, we explored for time bins of 50 ms, which electrodes show higher ERPs to intact than scrambled movies (Fig. 7b) to confirm observations from the difference curves of Fig. 6. Results confirmed that during the first 250 ms, parietal and occipital electrodes show stronger responses to the intact movies, in line with the earlier rise time. From 250 to 400 ms, no significant differences were observed. The pattern then reverses, with the scrambled condition triggering larger ERPs from 400 to 650 ms over parietal and occipital

electrodes. To provide an approximate mapping of the likely cortical sources for these differences at the electrode level, we performed a source localization of the ERP difference per time bin (Fig. 7b right columns). The similarity of this source distribution over time suggests that a similar network of brain regions, including bilateral parietal and right visual cortices, is responsible for the earlier rise-time and reduced amplitude in the intact compared to the scrambled condition. (Given the limitations of source localization, we do not provide a coordinate table for these localizations).

Fries (2015) and van Kerkoerle et al. (2014) report that within the visual system, feed-forward information and feed-back information across cortical regions increase coherence in the gamma band and the alpha/beta band, respectively. Taking stock from this observation, we compared the coherence between visual and supra-marginal ROIs (cf. Methods) across the intact and scrambled movies. If intact movies lead to more feedback predictions and less feedforward prediction errors, and the observations of Fries and van Kerkoerle et al. hold outside of the visual system, we would expect relatively more coherence in the alpha/beta band for intact, and more in the gamma band for scrambled movies. However, our analysis did not revealed any significant difference in coherence between these ROIs ( $q > 0.01$ ).



**Fig. 6.** ERP for the intact (red) and scrambled (blue) conditions. The inset below each ERP shows the difference curve (red for positive I-S and blue for negative) between intact and scrambled acts. The ERPs shown are also averaged across all the channels within the black shaded region shown on the sketch shown above each ERP. The channel groups enclose (a) OI1h, Oz, OI2h, O1, POO1, POO2, O2; (b) PO3, PPO1h, POz, PPO2h, PO4, P3, P1, Pz, P2, FP1, CPP5h, CPP3h, CPP1h, CPP2h, CPP4h, CPP6h; (c) CP3, CP1, CPz, CP2, CP4, CCP5h, CCP3h, CCP1h, CCP2h, CCP4h, CCP6h, C3, C1, Cz, C2, C4, FCC5h, FCC3h, FCC1h, FCC2h, FCC4h, FCC6h; (d) FC3, FC1, FCz, FC2, FC4, FFC3h, FFC1h, FFC2h, FFC4h, F3, F1, Fz, F2, F4; (e) AF3, AFF1h, AFF2h, AF4, AFP1, AFz, AFP2.

### 3.8. Intact movies lead to more $\mu$ -suppression at central electrodes

Our fMRI results suggest more consistent recruitment of the AOEN during intact movie processing than during the processing of scrambled movies. EEG power in the  $\mu$ -band at central electrodes C3, C4 and Cz is considered a proxy for AOEN recruitment (Arnstein et al., 2011; Pineda, 2005). During action observation and action execution, the power in the lower ( $\pm 10$  Hz) and upper ( $\pm 20$  Hz)  $\mu$ -band is reduced in comparison to control conditions (Pineda, 2005), and reduced  $\mu$ -power on a trial-to-trial basis co-occurs with higher BOLD signal in the AOEN (Arnstein et al., 2011). We therefore hypothesized that  $\mu$ -power at C3, C4 and Cz should be lower (more suppressed) during intact than scrambled movies, i.e., the power ratio  $\text{power}_{\text{intact}}/\text{power}_{\text{scrambled}}$  should fall below 1 in the  $\mu$ -range (10–20 Hz). Fig. 8 shows results that confirm this prediction. We found the power-ratio to be below 1 in the 10–20 Hz range in all the three electrodes over a range of time points relative to the camera change (Fig. 8 middle rows). Averaging power over time revealed significantly more  $\mu$ -suppression for intact movies than for scrambled movies (i.e. power-ratio  $< 1$ ,  $q_{\text{fdr}} < 0.05$ ) from 8 to 22 Hz for C3, 10–20 Hz for C4, and 12–28 Hz in Cz (Fig. 8 bottom row).

## 4. Discussion

In this experiment we set out to explore where and how the brain encodes sequence-level information when motor acts occur in sequences. Participants passively viewed a sequence of motor acts either in their natural, i.e., intact order or in a scrambled order. We measured their BOLD response using fMRI and their electrophysiological responses using EEG.

### 4.1. Mapping sequence level information using fMRI

Regions that show higher ISC during the processing of intact movies comprised three functional networks. One of these networks (“blue” network in Fig. 4d) became active early in watching the movies. This network consisted of the right visual cortices (including area V1, V2 and V3), and the left angular (PG) and dorsal premotor cortices often jointly associated with stimulus-driven spatial attention within the dorsal attention network (Nozawa et al., 2014; Nee and Jonides, 2014). The other two networks viz., the “red” and the “green” ones, were negatively correlated over time with the first network and became activated later when actions in the movies became more predictable. The red network in Fig. 4e was formed by bilateral parietal clusters including (a) the PF complex in the rostral inferior parietal lobe, (b) the primary somatosensory cortex including BA2,1,3a, and 3b, (c) the secondary somatosensory cortex including PV and SII, (d) the rostral intraparietal sulcus, (e) the left mid-dorsal insula, (f) right inferior (BA44), and (g) mid-frontal gyri. Much of this network became activated during execution of actions. This network demonstrates strong similarities with the AOEN (see below).

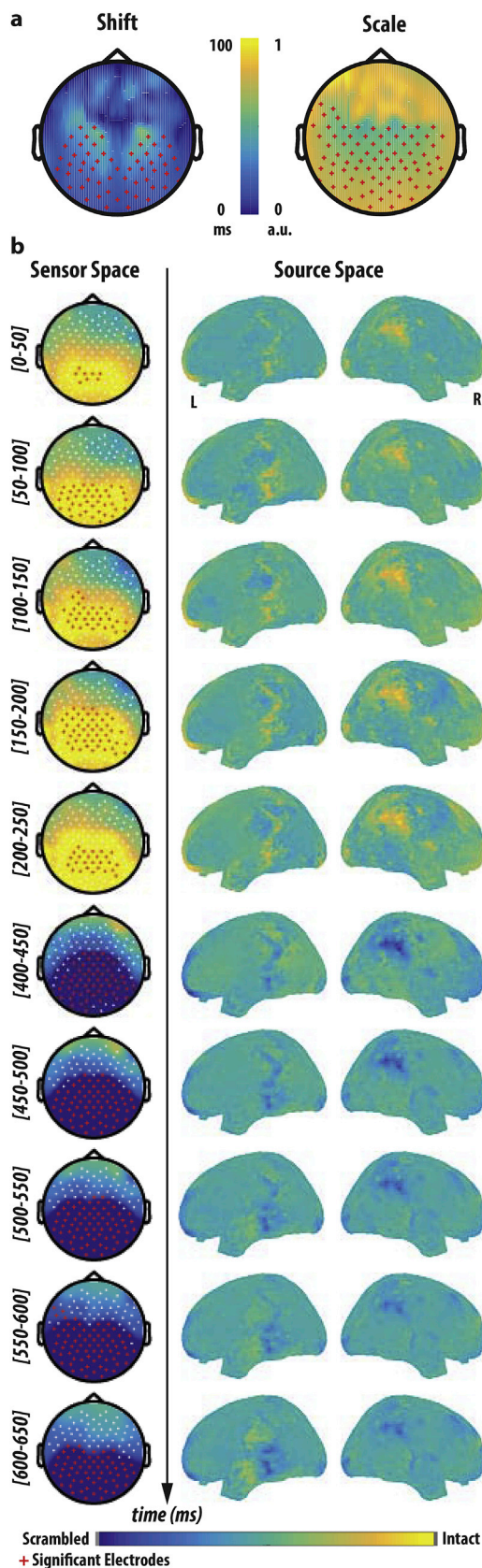
The green network consists of the superior parietal lobule bilaterally (including 7A/PC and 5L) and has often been associated with the integration of vision, somato-sensation, and action (Bremmer et al., 2001; Graziano and Cooke, 2006; Huang et al., 2012; Ishida et al., 2010; Schindler and Bartels, 2018).

We noted in the introduction that there is a difference of opinion in the neuroscience community as to whether information beyond single motor acts should fall within the AOEN or in the ToM network. We therefore wanted to investigate whether the three networks identified by us overlap with the AOEN and/or with the ToM network.

### 4.2. Overlap between sequence-level information and the AOEN

We found substantial overlap within the regions included in our AOEN localizer. Regions of significantly higher ISC during intact movies within the AOEN consisted of most of the red network, and, in particular the bilateral large parietal clusters including the PF complex, SI and SII, as well as smaller frontal clusters in the right ventral premotor cortex (BA44) and left mid-dorsal insula. AOEN overlapped also with the green network in the superior parietal areas BA5 and BA7. Given that the intact and scrambled movies contained identical segments showing individual motor acts, if the AOEN were only to represent individual motor acts in isolation, we would expect the ISC to be identical in the two conditions. That the ISC was higher for the intact movies shows that the brain activity in these regions is also sensitive to the transitions between actions, providing evidence that this network contains information at the sequence level beyond individual motor acts. The  $\mu$ -band in central EEG electrodes (C3, C4 and Cz), a putative proxy for AOEN activity (Arnstein et al., 2011; Pineda, 2005), was more suppressed during the intact than during the scrambled movies. This phenomenon is further evidence that there is greater recruitment of the AOEN in the intact case than in the case of scrambled movies.

Many of the regions of the red network in our human participants that overlap with the AOEN have homologues in the monkey brain that have been shown to contain mirror neurons. These include the human homologues of the following regions: (a) monkey premotor region F5 (Gallese et al., 1996; C Keyzers et al., 2003; Kohler et al., 2002; Umiltà et al., 2001), (b) monkey PF (Fogassi et al., 2005; Rozzi et al., 2008) and (c) monkey SII and adjacent sectors of SI (Hihara et al., 2015). This observation agrees with the predictions made by Hebbian learning models of the mirror neuron system that suggest that this system should encode the transitions between actions (Keyzers and Gazzola, 2014). It should be noted that any stimulus triggered brain activity must have two features for it to translate into an increase in ISC. It must occur at the same location across individuals after normalization and smoothing and it must take place at approximately the same time across participants (Supplementary Materials). The significant ISC increase we observe in the AOEN thus suggests that information about the transition between acts in the AOEN has these features – at least to some extent. This is in line with the



(caption on next column)

**Fig. 7. Differences in ERP across conditions.** (a) Topography of the parameters  $\lambda$  (left) and  $\alpha$  (right). Electrodes where  $\lambda$  differs significantly from zero (with ERPs from Intact movies rising faster than those from Scrambled movies) or  $\alpha$  differs significantly from one (with ERPs from Scrambled movies having higher magnitude) based on a cluster statistic are shown as red crosses. (b) Assessment of ERP differences as a function of time. On the left the topplot illustrates the difference in ERP between the intact and scrambled ERPs as a function of the time bin indicated. On the right side, a minimum norm source reconstruction of the difference. Blue indicates locations with Scrambled > Intact; yellow those were Intact > Scrambled. In all cases, sensors with significant differences between Scrambled and Intact (corrected for multiple comparison using a cluster statistic at  $p < 0.05$ ) are marked in red.

reliable timing observed in single mirror-neuron activity when observing single motor acts (Keysers et al., 2003; Kohler et al., 2002) and the consistency in the neural location activated by the sight of motor acts across individuals (Gazzola and Keysers, 2009).

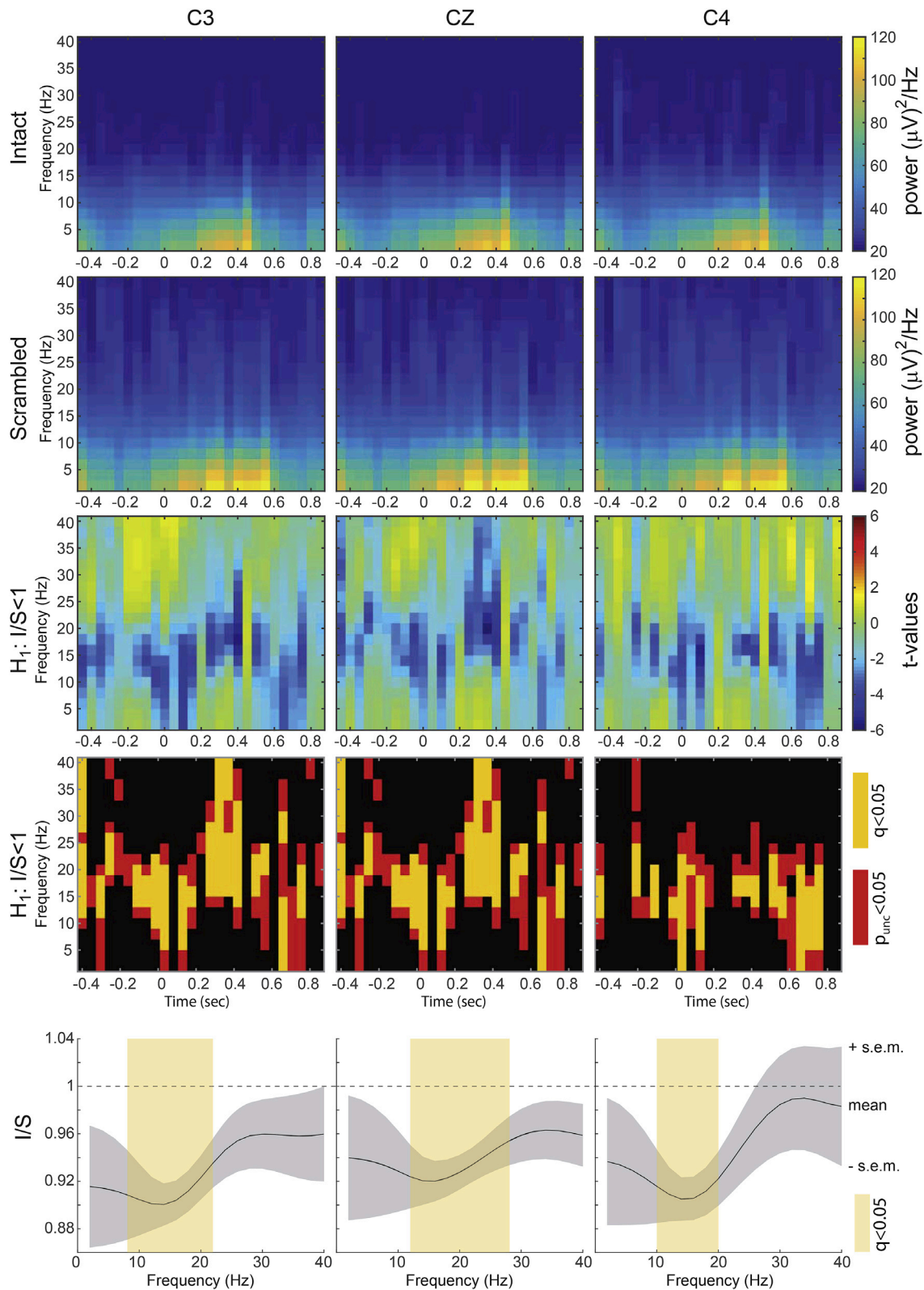
The sensitivity of AOEN regions to the sequence in which motor acts are observed supports the findings of a small number of studies that other situational information capable of predicting future acts can also modulate mirror neuron activity. Fogassi et al. (2005) have shown that if a monkey repeatedly witnesses someone grasping an object to bring it to the mouth, the activity of grasping mirror neurons is different from that in a context in which the monkey repeatedly sees someone grasp an object to place it in a container. Umiltà et al. (2001) have shown that when an occluding screen is placed in front of a graspable object, seeing a hand disappear behind this screen leads to a higher discharge in mirror neurons sensitive to the grasping action than seeing it disappear behind a screen that was placed in front of an empty platform. Finally, Iacoboni et al. (2005) showed that the ventral premotor node of the AOEN in humans responds differently to grasping a cup with a background of objects suggesting drinking vs. a background of objects suggesting washing the cup. Taken together, these findings suggest that the AOEN can take preceding acts, the objects in the scene and the current situation into account in generating its response to the sight of a particular motor act, thereby demonstrating its possession of information that goes beyond single motor acts. Further support for an association between the AOEN, the mirror neuron system and sequence-level information is obtained from the fact that most of the regions that carried sequence-level information in our analysis of action observation data were also activated when participants themselves manipulated objects (“\*” in Fig. 4) - a defining feature of mirror brain regions.

The green network also overlapped with the AOEN. Areas BA5 and BA7 are not typically considered part of the mirror neuron system. These regions have functional properties similar to those of monkey region VIP (Bremner et al., 2001; Huang et al., 2012; Schindler and Bartels, 2018). VIP neurons in monkey brains respond to (a) the haptic experience of touching an object with a specific body part, (b) the sight of objects close to and approaching that body part, and (c) the sight of another person being touched at the equivalent body part. VIP neurons, thus, can be said to provide a somatosensory analogue to traditional mirror neurons (Ishida et al., 2010). Neurons in this region have often been conceptualized as visually anticipating upcoming interactions between the body and objects (e.g. Graziano and Cooke, 2006). Their vicarious activation during observation of similar actions performed by others (Ishida et al., 2010) endows them with similar anticipatory functions during the observation of action. It is therefore perhaps not surprising, that the human homologues of VIP are sensitive to sequence information during action observation.

#### 4.3. Contributions of the ToM network

We did not observe any overlap between changes in ISC and the ToM network. However, a bGLM analysis that compared the average activity during the entire movie in the intact version to that in the scrambled version revealed that some voxels in the ToM network show higher





**Fig. 8. Higher mu-suppression for intact than scrambled movies.** Time frequency decompositions were performed for intact (top row) and scrambled (second row) movies relative to the camera-change ( $t = 0$ ) for the three central electrodes C3 (left), Cz (middle) and C4 (right column). A  $t$ -test comparing for each time point and frequency the power-ratio  $I/S$  against 1 revealed less power in the I than S condition (i.e. negative  $t$ -values) in the mu band (10–20 Hz) over multiple time-points (3rd row). Thresholding this comparison at  $p_{unc} < 0.05$  (red) or  $q_{fdr} < 0.05$  (yellow) reveals significant differences in the expected direction (i.e. ratio  $< 1$ ) over many time-points (4th row). The bottom row represents the power ratio obtained after averaging power over the entire time-window separately for each frequency and condition and then calculating the ratio (mean  $\pm$  sem across the 22 participants). The ratio  $I/S$  was then compared against 1 using a one-tailed  $t$ -test corrected for multiple comparison using  $q_{fdr} < 0.05$ . Time epochs of  $I/S < 1$  are then shown in yellow ( $q < 0.05$ ).

average activation in the scrambled condition though this did not translate into higher ISC in this network in that condition. As mentioned earlier (cf. Introduction), ISC requires activity to be stimulus-locked, i.e., to overlap in time across participants, while a bGLM does not (See Supplementary Excel Table for the difference between ISC and bGLM). This suggests that in contrast to the AOEN that becomes preferentially recruited in a stimulus-locked way when acts are embedded in natural sequences, the ToM network becomes preferentially recruited when motor acts deviate from natural sequences and then at times that vary across participants. The greater activity level noticed in this network in the scrambled actions case agrees well with the observation that ToM brain regions go online when witnessing implausible actions (Brass et al., 2007). That it is less stimulus-locked suggests that it reflects processes that are less automatic and more reflective, in line with the fact that parietal region PG is part of the default mode network that is associated with intrinsic rather than stimulus-driven cognition (Buckner et al., 2008). That different networks are recruited depending on whether sequences of acts are in natural or disturbed sequences suggests that the question whether sequence-level information is represented in the AOEN or in the ToM system is perhaps ill posed. Rather, these two systems are called up by sequence-level information under different circumstances (Keysers and Gazzola, 2007).

#### 4.4. Evidence for inhibition of predictable visual responses

The second aim of this project was to compare and contrast the predictions of different families of action observation models (Fig. 2). Our EEG data support inhibitory feedback models by showing that the amplitude and the latency of the brain's response to individual motor acts are less for the intact than for the scrambled versions. Source localization helps highlight the role of the parietal and visual cortices in generating these changes in timing and amplitude. On the basis of theoretical considerations alone, we had expected the *changes in amplitude* to be clearest over the visual cortex, and *changes in latencies* over parietal and premotor cortices – a separation that was not as evident in the EEG data which displayed both of these changes over occipital and parietal electrodes. The network identified in the source localization remained constant across time. This might either reflect the technical limitations of EEG in separating the activity of individual nodes or suggest that, in the sustained regime of activity induced by sequences of acts, the execution-observation network may begin to behave as one unified system rather than as a series of individual nodes with distinct temporal properties – a fact that is perhaps not surprising in a network with reciprocally connected nodes.

However, over the slower temporal scale of fMRI, the visual and parietal nodes of the AOEN had distinguishable time courses that caused them to fall within separable networks, with the parietal nodes peaking in activity later than the visual nodes. With regard to the distinction between inhibitory feedback and excitatory feedback between parietal and visual cortices, an anti-correlation is observed between these fMRI networks. This observation agrees with the inhibitory-feedback model according to which the neural representation of observed actions should shift from visual brain regions to parietal and premotor regions that generate increasingly constrained and more accurate predictions based on the somatosensory and motor connectivity pattern of the participant's own actions in the intact sequences (Keysers and Gazzola, 2014). The bGLM also shows that a number of brain regions associated with visual processing show reduced BOLD signals during the processing of the intact sequences – a finding that is compatible with the notion of visual cortices receiving inhibitory feed-back in the context of intact sequences. Other models would not predict such negative correlations. In the temporal domain, we find that responses in the intact sequences occurs about 50–100 ms earlier than in the scrambled situation (Fig. 7a). The order of magnitude of this temporal shift is roughly in line with what could be expected from the literature on perceptual momentum, according to which, when we see an intact sequence of motions, our brain anticipates

upcoming events tens of milliseconds in advance of their occurrence irrespective of whether they are inanimate objects (Freyd and Finke, 1984) or human actions (Verfaillie and Daems, 2002). Anticipation of this order of magnitude can also be observed in TMS experiments that measure motor facilitation to the vision of still frames that precede an action by ~100 ms (Urgesi et al., 2010), and is in line with what we had predicted in the theoretical accounts based on Hebbian learning (Keysers and Gazzola, 2014).

However, our data are limited in their ability to establish predictive coding within the AOEN. First, our measurements are unable to establish that the motor system is the cause for (i) the attenuation and acceleration of the responses in occipital electrodes in the EEG data for intact sequences or (ii) our ability to predict observed actions. Although it is compatible with such causal relationships, we need neuromodulation and/or lesion studies to establish such a causal connection. (Avenanti et al., 2017; Valchev et al., 2016). Secondly, a predictive coding framework posits that signals across visual and parietal cortices should provide *specific* predictions and prediction-errors about what action will come next, and how the observed action differs from these expectations – a specificity our data do not claim to have achieved yet. Designs in which we present sequences of actions that employ different body-parts (e.g. filling a glass with our hands, and then drinking it by ingesting it through the mouth) in combination with pattern classification (Etzel et al., 2008; Oosterhof et al., 2010) may be able to achieve greater specificity of these signals. Besides, predictive coding enables specific predictions about the direction of information flow across conditions. As fMRI is not a reliable source of information about the direction of information flow (Smith et al., 2011), we have not looked for such directionality in our fMRI data. The high temporal resolution of EEG would make it more suitable for such analyses. However, volume conduction in EEG contaminates signals across nearby sources, which is probably why our coherence analyses failed to provide significant differences in coherence between the conditions. Repeating similar experiments in patients with electro-corticographic (ECoG) electrodes could overcome this limitation. Finally, neither fMRI nor EEG data provide direct measures of neuronal activity. For fMRI, reduced BOLD activity in the intact case compared to the scrambled condition probably reflects reduced metabolic demands, but whether this is due to a reduction in spiking in neurons representing the expected action remains unclear due to the complicated relation between neural inhibition and metabolism (for a critical discussion see Mangia et al., 2009). EEG also does not measure the spiking of neurons, but the temporal summation of synchronized EPSP and IPSPs in pyramidal neurons (da Silva, 2010). The attenuation of the ERP in the intact sequence could thus reflect reduced firing in the pyramidal neurons providing the output of the visual cortex, as predicted by the inhibitory feed-back model, but could also represent a desynchronization of this activity without change in the number of spikes. Experiments in which the activity of well-characterized neurons in the primate visual cortex is measured as the subjects observe both intact and scrambled sequences could help disambiguate some of the results of our non-invasive measurements reported here.

How the AOEN learns to encode information about transitions between acts is an interesting question albeit one that we cannot answer at the present stage of our work. Let us imagine an agent performing a sequence of acts A, B, and C. We know that systematic delays would occur between motor commands and sensory reafferences. When premotor or parietal regions trigger a motor command, there is a ~100 ms gap before the body part executes the action (Graziano et al., 2005), and another ~100 ms elapses before sensory information (visual, acoustic and somatosensory) travels back through sensory cortices to high-level cortices (Keysers et al., 2001, 2003; Kohler et al., 2002). Hence, in the parietal cortex, when motor commands for act B are triggered, sensory re-afferent information about act A would still be encoded in the synaptic input due to the ~200 ms sensory-motor delays. This way, Hebbian synaptic learning would reinforce synaptic connections between sensory information about act A and motor commands about act B (Keysers and Gazzola, 2014). Given the way somatosensory and motor cortices are

functionally interconnected during action observation (Valchev et al., 2016), we could conceive of sensory-motor loops that would represent the sequence of A->B, then from B->C, etc., based on the statistics of our own past actions. Studies of synaptic plasticity in animal models could test this hypothesis, and studies that vary the statistics of a participant's past action-transition-probabilities could test whether past motor experiences are indeed a significant source of these predictive signals.

## 5. Conclusion

Our data suggest that when acts are arranged in sequences, the additional information is represented in separable networks depending on whether the sequences adhere or do not adhere to the statistics of natural actions. When acts adhere to the natural statistics of our own actions, regions overlapping with the AOEN encode sequence-level information in a stimulus-locked fashion. An inhibitory feedback architecture, i.e., one in which the visual processing of acts is inhibited by expectations derived from previous actions is then most compatible with the pattern of responses we observed: (a) a reduction of visually evoked responses in EEG, (b) a reduction in average BOLD activity in the visual cortex for intact compared to scrambled sequences and (c) a negative correlation between BOLD activity in the visual region and that in the AOEN nodes.

When the order of the acts violates the natural statistics, brain regions associated with ToM seem to encode sequence-level information in a spatially consistent but temporally more variable way. This is seen in an increase of average BOLD activity without changes in ISC in this network. These findings call for more in-depth studies using brain activity manipulation and investigating lesions that could throw fresh light on whether sequence processing requires the AOEN when sequences fit natural statistics and ToM regions when they violate such statistics.

## Declarations of interest

None.

## Acknowledgments

This work was supported by the Netherlands Organisation for Scientific Research (VENI: 451-09-006, VIDI: 452-14-015 to V.G., Brain and Cognition: 433-09-253), the Brain and Behavior Research Foundation (NARSAD young investigator 22453 to V.G.) and the European Research Council of the European Commission (ERC-StG-312511 to C.K.). We thank Judith Suttrup for acquiring and analysing the data necessary to define the mirror network mask, Henk Stoffels and Abelraham Abdelgabar for helping with the stimulus recording, and Juha Pajula for help with using the ISC toolbox. We thank Michael Spezio for helpful discussions on technical, conceptual and statistical aspects of the work, and Uri Hasson for helpful discussion on ISC data analysis and interpretation. We thank Dr Thomas Chacko for helping us improve the clarity and readability of the text. Correspondence should be addressed to v.gazzola@nin.knaw.nl or c.keyzers@nin.knaw.nl.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.neuroimage.2018.08.056>.

## References

- Arnstein, D., Cui, F., Keyzers, C., Maurits, N.M.N.M., Gazzola, V., 2011.  $\mu$ -suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. *J. Neurosci.* 31, 14243–14249. <https://doi.org/10.1523/JNEUROSCI.0963-11.2011>.
- Avenanti, A., Candidi, M., Urgesi, C., 2013. Vicarious motor activation during action perception: beyond correlational evidence. *Front. Hum. Neurosci.* 7. <https://doi.org/10.3389/fnhum.2013.00185>.
- Avenanti, A., Paracampo, R., Anella, L., Tidoni, E., Aglioti, S.M., 2017. Boosting and decreasing action prediction abilities through excitatory and inhibitory tDCS of inferior frontal cortex. *Cerebellum. Cortex* 1–15. <https://doi.org/10.1093/cercor/bhx041>.
- Brass, M., Schmitt, R.M., Spengler, S., Gergely, G., 2007. Investigating action understanding: inferential processes versus action simulation. *Curr. Biol.* 17, 2117–2121. <https://doi.org/10.1016/j.cub.2007.11.057>.
- Bremner, F., Schlack, A., Shah, N.J., Zafiris, O., Kubischik, M., Hoffmann, K., Zilles, K., Fink, G.R., 2001. Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron* 29, 287–296.
- Buccino, G., Vogt, S., Ritzl, A., Fink, G.R., Zilles, K., Freund, H.-J., Rizzolatti, G., 2004. Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron* 42, 323–334.
- Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. <https://doi.org/10.1196/annals.1440.011>.
- Caramazza, A., Anzellotti, S., Strnad, L., Lingnau, A., 2014. Embodied cognition and mirror neurons: a critical assessment. *Annu. Rev. Neurosci.* 37, 1–15. <https://doi.org/10.1146/annurev-neuro-071013-013950>.
- Caspers, S., Zilles, K., Laird, A.R., Eickhoff, S.B., 2010. ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* 50, 1148–1167. <https://doi.org/10.1016/j.neuroimage.2009.12.112>.
- Cisek, P., Kalaska, J.F., 2004. Neural correlates of mental rehearsal in dorsal premotor cortex. *Nature* 431, 993–996. <https://doi.org/10.1038/nature03005>.
- Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67.
- da Silva, F.L., 2010. EEG: origin and measurement. In: Mulert, C., Lemieux, L. (Eds.), *EEG - FMRI: Physiological Basis, Technique, and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 19–38. [https://doi.org/10.1007/978-3-540-87919-0\\_2](https://doi.org/10.1007/978-3-540-87919-0_2).
- de Wit, M.M., Buxbaum, L.J., 2017. Critical motor involvement in prediction of human and non-biological motion trajectories. *J. Int. Neuropsychol. Soc.* 23, 171–184. <https://doi.org/10.1017/S1355617716001144>.
- Dinstein, I., Hasson, U., Rubin, N., Heeger, D.J., 2007. Brain areas selective for both observed and executed movements. *J. Neurophysiol.* 98, 1415–1427. <https://doi.org/10.1152/jn.00238.2007>.
- Disbrow, E., Litinas, E., Recanzone, G.H., Padberg, J., Krubitzer, L., 2003. Cortical connections of the second somatosensory area and the parietal ventral area in macaque monkeys. *J. Comp. Neurol.* 462, 382–399. <https://doi.org/10.1002/cne.10731>.
- Dushanova, J., Donoghue, J., 2010. Neurons in primary motor cortex engaged during action observation. *Eur. J. Neurosci.* 31, 386–398. <https://doi.org/10.1111/j.1460-9568.2009.07067.x>.
- Etzel, J.A., Gazzola, V., Keyzers, C., 2008. Testing simulation theory with cross-modal multivariate classification of fMRI data. *PLoS One* 3, e3690. <https://doi.org/10.1371/journal.pone.0003690>.
- Fiebach, C.J., Schubotz, R.I., 2006. Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? *Cortex* 42, 499–502.
- Filimon, F., Nelson, J.D., Hagler, D.J., Sereno, M.I., 2007. Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage* 37, 1315–1328. <https://doi.org/10.1016/j.neuroimage.2007.06.008>.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667. <https://doi.org/10.1126/science.1106138>.
- Freyd, J.J., Fiske, R.A., 1984. Representational momentum. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 126–132.
- Fries, P., 2015. Rhythms for cognition: communication through coherence. *Neuron* 88, 220–235. <https://doi.org/10.1016/j.neuron.2015.09.034>.
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. *Brain* 119 (Pt 2), 593–609. <https://doi.org/10.1093/brain/119.2.593>.
- Gallese, V., Keyzers, C., Rizzolatti, G., 2004. A unifying view of the basis of social cognition. *Trends Cognit. Sci.* 8, 396–403. <https://doi.org/10.1016/j.tics.2004.07.002>.
- Gazzola, V., Keyzers, C., 2009. The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebr. Cortex* 19, 1239–1255. <https://doi.org/10.1093/cercor/bhn181>.
- Grafton, S.T., Hamilton, A.F. de C., 2007. Evidence for a distributed hierarchy of action representation in the brain. *Hum. Mov. Sci.* 26, 590–616. <https://doi.org/10.1016/j.humov.2007.05.009>.
- Graziano, M.S.A., Aflalo, T.N.S., Cooke, D.F., 2005. Arm movements evoked by electrical stimulation in the motor cortex of monkeys. *J. Neurophysiol.* 94, 4209–4223. <https://doi.org/10.1152/jn.01303.2004>.
- Graziano, M.S.A., Cooke, D.F., 2006. Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia* 44, 2621–2635.
- Grèzes, J., Armony, J.L., Rowe, J., Passingham, R.E., 2003. Activations related to “mirror” and “canonical” neurones in the human brain: an fMRI study. *Neuroimage* 18, 928–937.
- Grosbras, M.-H., Beaton, S., Eickhoff, S.B., 2012. Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. *Hum. Brain Mapp.* 33, 431–454. <https://doi.org/10.1002/hbm.21222>.
- Hasson, U., Ghazanfar, A.A., Galantucci, B., Garrod, S., Keyzers, C., 2012. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cognit. Sci.* 16, 114–121. <https://doi.org/10.1016/j.tics.2011.12.007>.



- Hasson, U., Malach, R., Heeger, D.J., 2010. Reliability of cortical activity during natural stimulation. *Trends Cognit. Sci.* 14, 40–48. <https://doi.org/10.1016/j.tics.2009.10.011>.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. *Science* 303 (80), 1634–1640. <https://doi.org/10.1126/science.1089506>.
- Hihara, S., Taoka, M., Tanaka, M., Iriki, A., 2015. Visual responsiveness of neurons in the secondary somatosensory area and its surrounding parietal operculum regions in awake macaque monkeys. *Cerebr. Cortex* 25, 4535–4550. <https://doi.org/10.1093/cercor/bhv095>.
- Huang, R.-S., Chen, C.-f., Tran, A.T., Holstein, K.L., Sereno, M.I., 2012. Mapping multisensory parietal face and body areas in humans. *Proc. Natl. Acad. Sci. Unit. States Am.* 109, 18114–18119. <https://doi.org/10.1073/pnas.1207946109>.
- Iacoboni, M., Koski, L.M., Brass, M., Bekkering, H., Woods, R.P., Dubeau, M.C., Mazziotta, J.C., Rizzolatti, G., 2001. Reafferent copies of imitated actions in the right superior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13995–13999. <https://doi.org/10.1073/pnas.241474598>.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., Rizzolatti, G., 2005. Grasping the intentions of others with One's own mirror neuron system. *PLoS Biol.* 3, e79. <https://doi.org/10.1371/journal.pbio.0030079>.
- Ishida, H., Nakajima, K., Inase, M., Murata, A., 2010. Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *J. Cognit. Neurosci.* 22, 83–96. <https://doi.org/10.1162/jocn.2009.21185>.
- Kaufman, L., Rousseeuw, P.J., 1990. *Finding Groups in Data*. John Wiley & Sons, New York.
- Kauppi, J.-P., Jääskeläinen, I.P., Sams, M., Tohka, J., 2010. Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency. *Front. Neuroinf.* 4, 5. <https://doi.org/10.3389/fninf.2010.00005>.
- Kauppi, J.-P., Pajula, J., Tohka, J., 2014. A versatile software package for inter-subject correlation based analyses of fMRI. *Front. Neuroinf.* 8, 2. <https://doi.org/10.3389/fninf.2014.00002>.
- Keysers, C., Gazzola, V., 2014. Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130175–20130175. <https://doi.org/10.1098/rstb.2013.0175>.
- Keysers, C., Gazzola, V., 2007. Integrating simulation and theory of mind: from self to social cognition. *Trends Cognit. Sci.* 11, 194–196. <https://doi.org/10.1016/j.tics.2007.02.002>.
- Keysers, C., Kaas, J.H., Gazzola, V., 2010. Somatosensation in social perception. *Nat. Rev. Neurosci.* 11, 417–428. <https://doi.org/10.1038/nrn2833>.
- Keysers, C., Kohler, E., Umiltà, M.A., Nanetti, L., Fogassi, L., Gallese, V., 2003. Audiovisual mirror neurons and action recognition. In: *Experimental Brain Research*, pp. 628–636. <https://doi.org/10.1007/s00221-003-1603-5>.
- Keysers, C., Paracampo, R., Gazzola, V., 2018. What Neuromodulation and Lesion studies tell us about the function of the mirror neuron system and embodied cognition. *Curr. Opin. Psychol.*
- Keysers, C., Perrett, D.I., 2004. Demystifying social cognition: a Hebbian perspective. *Trends Cognit. Sci.* 8, 501–507. <https://doi.org/10.1016/j.tics.2004.09.005>.
- Keysers, C., Xiao, D.K., Földiák, P., Perrett, D.I., 2001. The speed of sight. *J. Cognit. Neurosci.* 13, 90–101.
- Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. *Cognit. Process.* 8, 159–166. <https://doi.org/10.1007/s10339-007-0170-2>.
- Kilner, J.M., Frith, C.D., 2008. Action observation: inferring intentions without mirror neurons. *Curr. Biol.* 18, R32–R33. <https://doi.org/10.1016/j.cub.2007.11.008>.
- Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848. <https://doi.org/10.1126/science.1070311>.
- Kraskov, A., Philipp, R., Waldert, S., Vigneswaran, G., Quallo, M.M., Lemon, R.N., 2014. Corticospinal mirror neurons. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130174–20130174. <https://doi.org/10.1098/rstb.2013.0174>.
- Lerner, Y., Honey, C.J., Silbert, L.J., Hasson, U., 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 31, 2906–2915. <https://doi.org/10.1523/JNEUROSCI.3684-10.2011>.
- Lewis, J.W., Van Essen, D.C., 2000. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–137.
- Makris, S., Urgesi, C., 2015. Neural underpinnings of superior action prediction abilities in soccer players. *Soc. Cognit. Affect Neurosci.* 10, 342–351. <https://doi.org/10.1093/scan/nsu052>.
- Mangia, S., Giove, F., Tkáč, I., Logothetis, N.K., Henry, P.-G., Olman, C.A., Maraviglia, B., Di Salle, F., Ugurbil, K., 2009. Metabolic and hemodynamic events after changes in neuronal activity: current hypotheses, theoretical predictions and in vivo NMR experimental findings. *J. Cerebr. Blood Flow Metabol.* 29, 441–463. <https://doi.org/10.1038/jcbfm.2008.134>.
- Mar, R.A., 2011. The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol.* 62, 103–134. <https://doi.org/10.1146/annurev-psych-120709-145406>.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Meth.* 164, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>.
- Maunsell, J.H., van Essen, D.C., 1983. The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J. Neurosci.* 3, 2563–2586.
- Molenberghs, P., Cunnington, R., Mattingley, J.B., 2012. Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neurosci. Biobehav. Rev.* 36, 341–349. <https://doi.org/10.1016/j.neubiorev.2011.07.004>.
- Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., Orban, G.A., 2011. Action observation circuits in the macaque monkey cortex. *J. Neurosci.* 31, 3743–3756. <https://doi.org/10.1523/JNEUROSCI.4803-10.2011>.
- Nee, D.E., Jonides, J., 2014. Frontal-medial temporal interactions mediate transitions among representational states in short-term memory. *J. Neurosci.* 34, 7964–7975. <https://doi.org/10.1523/JNEUROSCI.0130-14.2014>.
- Nozawa, T., Sugiura, M., Yokoyama, R., Ihara, M., Kotozaki, Y., Miyauchi, C.M., Kanno, A., Kawashima, R., 2014. Ongoing activity in temporally coherent networks predicts intra-subject fluctuation of response time to sporadic executive control demands. *PLoS One* 9, e99166. <https://doi.org/10.1371/journal.pone.0099166>.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Oosterhof, N.N., Wiggett, A.J., Diedrichsen, J., Tipper, S.P., Downing, P.E., 2010. Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *J. Neurophysiol.* 104, 1077–1089. <https://doi.org/10.1152/jn.00326.2010>.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869. <https://doi.org/10.1155/2011/156869>.
- Pineda, J.A., 2005. The functional significance of mu rhythms: translating “seeing” and “hearing” into “doing”. *Brain Res. Brain Res. Rev.* 50, 57–68. <https://doi.org/10.1016/j.brainresrev.2005.04.005>.
- Pons, T.P., Kaas, J.H., 1986. Corticocortical connections of area 2 of somatosensory cortex in macaque monkeys: a correlative anatomical and electrophysiological study. *J. Comp. Neurol.* 248, 313–335. <https://doi.org/10.1002/cne.902480303>.
- Rizzolatti, G., Sinigaglia, C., 2010. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat. Rev. Neurosci.* 11, 264–274. <https://doi.org/10.1038/nrn2805>.
- Rozzi, S., Calzavara, R., Belmalih, A., Borra, E., Gregoriou, G.G., Matelli, M., Luppino, G., 2006. Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cerebr. Cortex* 16, 1389–1417. <https://doi.org/10.1093/cercor/bhj076>.
- Rozzi, S., Ferrari, P.F., Bonini, L., Rizzolatti, G., Fogassi, L., 2008. Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *Eur. J. Neurosci.* 28, 1569–1588. <https://doi.org/10.1111/j.1460-9568.2008.06395.x>.
- Schindler, A., Bartels, A., 2018. Integration of visual and non-visual self-motion cues during voluntary head movements in the human brain. *Neuroimage* 172, 597–607. <https://doi.org/10.1016/j.neuroimage.2018.02.006>.
- Schubotz, R.L., Sakreida, K., Tittgemeyer, M., von Cramon, D.Y., 2004. Motor areas beyond motor performance: deficits in serial prediction following ventrolateral premotor lesions. *Neuropsychology* 18, 638–645. <https://doi.org/10.1037/0894-4105.18.4.638>.
- Schubotz, R.L., von Cramon, D.Y., 2001. Interval and ordinal properties of sequences are associated with distinct premotor areas. *Cerebr. Cortex* 11, 210–222.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34. <https://doi.org/10.1016/j.neubiorev.2014.01.009>.
- Simony, E., Honey, C.J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., Hasson, U., 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* 7, 12141. <https://doi.org/10.1038/ncomms12141>.
- Simos, P.G., Kavroulakis, E., Maris, T., Papadaki, E., Boursianis, T., Kalatzakis, G., Savaki, H.E., 2017. Neural foundations of overt and covert actions. *Neuroimage* 152, 482–496. <https://doi.org/10.1016/j.neuroimage.2017.03.036>.
- Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13040–13045. <https://doi.org/10.1073/pnas.0905267106>.
- Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for FMRI. *Neuroimage* 54, 875–891. <https://doi.org/10.1016/j.neuroimage.2010.08.063>.
- Stephens, G.J., Honey, C.J., Hasson, U., 2013. A place for time: the spatiotemporal structure of neural dynamics during natural audition. *J. Neurophysiol.* 110, 2019–2026. <https://doi.org/10.1152/jn.00268.2013>.
- Thioux, M., Gazzola, V., Keysers, C., 2008. Action understanding: how, what and why. *Curr. Biol.* 18, R431–R434. <https://doi.org/10.1016/j.cub.2008.03.018>.
- Tkach, D., Reimer, J., Hatsopoulos, N.G., 2007. Congruent activity during action and action observation in motor cortex. *J. Neurosci.* 27, 13241–13250. <https://doi.org/10.1523/JNEUROSCI.2895-07.2007>.
- Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., Rizzolatti, G., 2001. I know what you are doing: a neurophysiological study. *Neuron* 31, 155–165.
- Urgesi, C., Candidi, M., Avenanti, A., 2014. Neuroanatomical substrates of action perception and understanding: an anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Front. Hum. Neurosci.* 8. <https://doi.org/10.3389/fnhum.2014.00344>.
- Urgesi, C., Maieron, M., Avenanti, A., Tidoni, E., Fabbro, F., Aglioti, S.M., 2010. Simulating the future of actions in the human corticospinal system. *Cerebr. Cortex* 20, 2511–2521. <https://doi.org/10.1093/cercor/bhp292>.
- Valchev, N., Gazzola, V., Avenanti, A., Keysers, C., 2016. Primary somatosensory contribution to action observation brain activity-combining fMRI and cTBS. *Soc. Cognit. Affect Neurosci.* 11, nsw029. <https://doi.org/10.1093/scan/nsw029>.
- van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., Roelfsema, P.R., 2014. Alpha and gamma oscillations characterize feedback

- and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. Unit. States Am.* 111, 14332–14341. <https://doi.org/10.1073/pnas.1402773111>.
- Verfaillie, K., Daems, A., 2002. Representing and anticipating human actions in vision. *Vis. cogn.* 9, 217–232. <https://doi.org/10.1080/13506280143000403>.
- Vigneswaran, G., Philipp, R., Lemon, R.N., Kraskov, A., 2013. M1 corticospinal mirror neurons and their role in movement suppression during action observation. *Curr. Biol.* 23, 236–243. <https://doi.org/10.1016/j.cub.2012.12.006>.