# Predictive neural representations of naturalistic dynamic input

Short title: Dynamic representational similarity analysis

Ingmar E.J. de Vries,[1]* Moritz F. Wurm[1]

[1] Centre for Mind/Brain Sciences (CIMeC), University of Trento, 38068 Rovereto, Italy.

* i.e.j.de.vries@gmail.com

## Abstract

Our capacity to interact with our dynamic world in a timely manner (e.g., catch a ball) suggests that our brain generates predictions of unfolding external dynamics. While theories assume such neural predictions, empirical evidence typically only captures a snapshot or indirect consequence of prediction, and uses simple static stimuli of which predictability is manipulated. However, the rich dynamics of predictive representations remain largely unexplored. We present a novel dynamic extension to representational similarity analysis (RSA) that uses temporally variable models to capture naturalistic dynamic stimuli, and demonstrate both lagged and predictive neural representations in source-reconstructed MEG data. Interestingly, the predictive representations show a hierarchical pattern, such that higher-level stimulus features are predicted earlier in time, while lower-level features are predicted closer in time to the actual sensory input. This promising new approach opens the door for addressing important outstanding questions on how our brain represents and predicts our dynamic world.

# Introduction

To successfully navigate dynamic environments, such as traffic, or interacting with others in general, our brain needs to continuously update its representations of dynamic external information. For example, when playing football, we require an updated representation of the position and motion of other players, and the ball. This poses a fundamental problem: how does the brain cope with a stream of dynamic input? It takes time to transmit and process information along the hierarchy of the visual system. Nevertheless, we do not perceive the world in a lagged manner. Instead, our capacity to interact with dynamic stimuli in a timely manner demonstrates that our brain predicts future events. Without such predictions, there would be a substantial delay between external states, our perception of these states, and our subsequent reactions. Indeed, contemporary theories describe our brain as a predictive organ[1–5]. They postulate that our brain uses previous experiences to build hypotheses about future states of the world, which are continuously updated based on new evidence, acquired by sampling the world through our senses. The predictive nature of our brain has indeed been demonstrated empirically throughout the sensory processing hierarchy; e.g., for the prediction of future object position[6,7] or orientation[8,9] in visual cortex, an earlier neural representation of stimulus position when its apparent motion is predictable[10], a biasing of motion direction representation in early sensory areas towards the predicted direction[11], increased top-down functional connectivity during observation of predictable action sequences[12], a context-based predictive neural representation of words during naturalistic reading[13], or a predictive representation of expected tone pitch in auditory cortex[14]. Additionally, ample research has demonstrated specific brain responses to violations of prediction (i.e., prediction errors); e.g., to an unexpected face in a learned sequence in the macaque face processing network[15], to an human-like android exhibiting unexpectedly mechanical movement kinematics[16,17], or to an unexpected tone occurring in an learned sequence[18,19]. While thus providing clear evidence for neural prediction, empirical research typically looks at a 'snapshot' or consequence of a prediction, and uses simple, often static, stimuli of which the predictability is manipulated. However, we live in a dynamic world that necessitates us to continuously update our predictions based on a combination of prior knowledge and the integration of constantly changing information that our environment affords. Previous paradigms and analysis approaches precluded exploring the rich temporal dynamics of predictive representations in naturalistic settings, in particular across hierarchical levels of processing.

Here we present a novel dynamic extension to representational similarity analysis (RSA) to investigate how our brain predicts future sensory input. Conventional RSA typically uses feature models of stimuli at different levels of processing (e.g., shape, category) to investigate how these features are represented in the brain[20]. For RSA, subjects are typically presented with a set of stimuli, and neural and model stimulus representations are quantitatively related to infer whether information captured by a model is present in the neural signal. Neural and model stimulus representations are operationalized by representational dissimilarity matrices (RDMs), in which each entry represents the dissimilarity (i.e., distance) between a pair of stimuli in a (computational) model description of stimulus features (e.g., shape or category), or the dissimilarity in the pattern of the brain response to these stimuli (e.g., pattern over functional magnetic resonance imaging (fMRI) voxels or magnetoencephalography (MEG) sensors). At a final step, RSA quantifies how well the model RDM captures the representational variance in the neural RDM. Over the past decade, RSA has been a major influential advancement in cognitive/systems neuroscience. It has provided crucial insights about where (using fMRI[20]), and when (MEG[21]) our brain represents different stimulus features. RSA applied to MEG typically uses static stimuli and models to uncover the temporal cascade of stimulus processing from low-level features (e.g., shape) in early visual cortex, to later processing of high-level features (e.g., category) in higher brain areas. While invaluable for our understanding of sensory processing, static stimuli and models preclude investigating how our brain predicts future sensory input in dynamic environments. We present a novel dynamic extension to RSA (hereafter dynamic RSA or dRSA; see Fig. 1 for an explanation) that uses temporally variable models (i.e., a unique model RDM for each time point), to capture the representational nature of dynamic stimuli (e.g., action sequences), along various dimensions of interest (e.g., from low-level visual to viewpoint independent body posture and motion).

We applied this approach to source-reconstructed MEG data of healthy human subjects who observed videos of complex ballet dancing sequences, and demonstrate that dRSA robustly reveals predictive and lagged neural representations of naturalistic dynamic input. Interestingly, in line with the hierarchical nature of predictive processing[22,23], the predictive representations observed here also show a hierarchical pattern, such that more complex and abstract stimulus features are predicted earlier in time, while lower-level features are predicted closer in time to the actual sensory input. This new approach offers a window into predictive processing in naturalistic dynamic settings that we encounter in our everyday life, and can be extended to any context or modality (e.g., music, naturalistic soundscapes, (sign) language), and any signal or organism (e.g., EEG, EMG, NIRS, EcoG, single cell recordings). As such, it offers a novel approach for answering important questions regarding how we process and predict the dynamic world we live in.

# Results

## Procedure

We investigated when our brain represents the dynamics of the world at different hierarchical levels of visual processing. Specifically, we developed a novel dynamic extension to RSA that uses temporally variable models to capture neural representations of time-varying stimuli (Fig. 1). We applied this approach to MEG data of 22 healthy human subjects who observed ~36 repetitions of 14 unique 5-sec ballet dancing videos (Table S1 and Fig. S1a). Attention to the ballet dancer

was confirmed by high performance on occasional catch trials (22% of all trials) in which the video stopped unexpectedly, and subjects were asked to (dis)confirm a statement regarding the ballet dancer's motion (Fig. S1b and c). Similarly, eye fixation was ensured by an occasional subtle change in color of the fixation cross that was displayed in the center, to which subjects responded successfully (Fig S1b and c). We characterized the stimuli at different hierarchical levels of complexity by means of several stimulus models (i.e., model RDMs; Fig. 1b and e), describing the stimuli in terms of pixelwise luminance values, optical flow vectors (magnitude and direction), absolute and relative (i.e., viewpoint invariant) body posture, and absolute and relative body motion (Fig. S2). Body posture and motion models were based on 3D recordings of 13 kinematic markers placed on the ballet dancer. MEG data were source-reconstructed and 6 anatomical ROIs (Fig. S3) were selected, namely three visual areas (V1, V2, and V3+4) and three nodes of the action observation network (AON); the lateral occipitotemporal cortex (LOTC), the anterior inferior parietal lobe (aIPL), and the ventral premotor cortex (PMv), to create neural representations (i.e., neural RDMs; Fig. 1e).

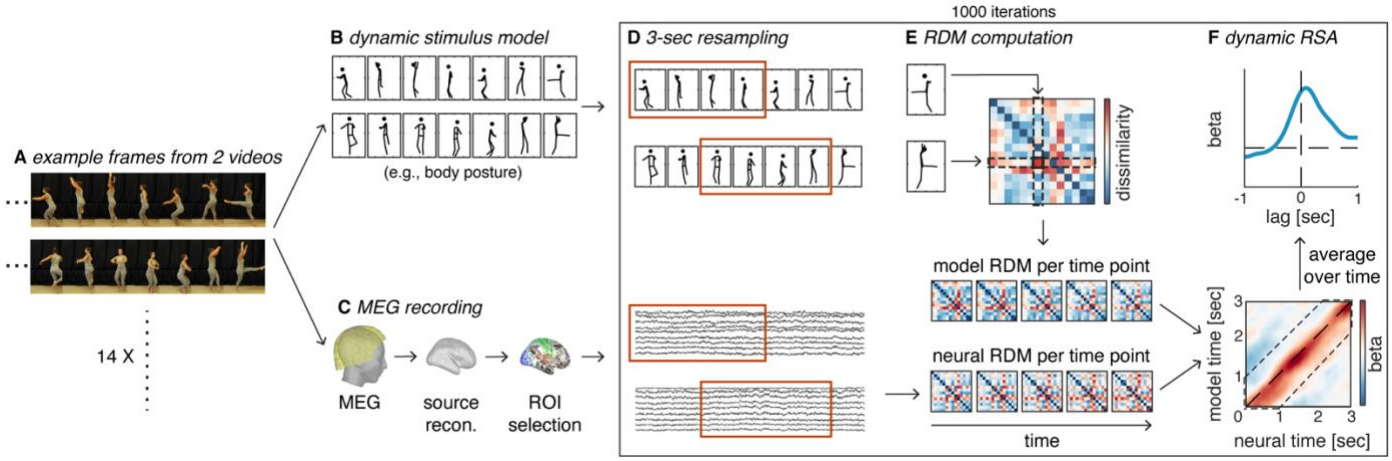


**Figure 1. Dynamic RSA.** See 'Materials and Methods' for details. (a) Subjects observed 14 unique 5-sec ballet dancing videos in the MEG scanner. (b) Various stimulus models are created to capture different stimulus features at each time point such as body posture, body motion, etc. (c) MEG activity is MNE source-reconstructed, and signals from vertices within an ROI are used as features for the next step. (d) Across 1000 iterations, a unique 3-sec window is randomly selected for each of the 14 unique stimuli. For a given stimulus, the time window is the same for the neural data and the stimulus model. (e) Subsequently, on each resampling iteration, a neural and model RDM are created based on the pairwise dissimilarity in the neural response to each of the 14 stimuli and the pairwise dissimilarity in the stimulus model, respectively. These neural and model RDMs are dynamic as they are computed for each time sample within the 3-sec time window (here shown for 5 example time points). (f) Last, similarity between neural and model RDMs is computed for each neural-by-model time point, resulting in a 2-dimensional neural-time by model-time dynamic RSA matrix, which is subsequently averaged over the 1000 resampling iterations (lower panel). The dRSA matrix can be further averaged over stimulus time for each lag to create a lag-plot (i.e., the lag between neural and model RDM; upper panel).

In dynamic RSA, similarity between neural and model RDMs is computed at all neural-by-model time points, resulting in a neural-time by model-time dynamic RSA plot (Fig. 1f; lower panel). Hereby, dRSA allows investigating the match between the model at a given time point and the neural representation at the same or different time points (i.e., earlier or later). In case of pure bottom-up processing, one expects a lag between a time point-specific visual model RDM and the best-matching neural RDM in V1 (i.e., the time needed for information to pass from the retina to V1). Such a lagged representation would be reflected by high similarity shifted away from the diagonal into the lower right triangle of the dRSA plot in Figure 1f (lower panel), or similarly, after averaging over stimulus time, a peak to the right of the zero-lag vertical midline in the lag-plot in Figure 1f (upper panel). In contrast, a negative lag should be observed for predictive neural representations (i.e., higher similarity in the upper left triangle of the dRSA plot or a peak on the left side in the lag-plot), in which case neural representational content predicts future model states. We adopted principal component regression (PCR) weights as measure of similarity between the neural and model RDMs, in order to test whether a specific model RDM could explain variance in the neural RDM, while regressing out other covarying model RDMs, and confirmed effectiveness of this approach through simulations (see "*Materials and Methods"* and Fig. S4). To control for potential eye movement artifacts, we additionally regressed out an RDM based on the individual-subject eye position as recorded with an eye-tracker. Furthermore, note that arbitrary stimulus-specific temporal inhomogeneities (e.g., video onset, motion onsets and offsets, predictability, model RDM variance), would be reflected in the representational dynamics. To attenuate this, across 1000 iterations we resampled the 5-sec stimuli to a random 3-sec window uniquely selected for each of the 14 unique stimuli (Fig. 1d), computed dRSA as described above, and averaged across iterations.

**Predictive neural representation of future body motion**

Most importantly, dRSA revealed that the ballet dancer's body motion was represented in a broad temporal window preceding the actual input by ~640–430 msec for relative body motion and ~300–150 msec for absolute body motion (Fig. 2a, 3a and S5), indicating that these neural representations predicted future body motion. The predictive representation of motion was further confirmed by a computer vision model specifically describing the direction (but not magnitude) of apparent motion for each pixel in de video in terms of optical flow vectors (~90 msec). Interestingly, the latency of prediction seemed related to the hierarchical level of the motion representation, with the largest predictive lag for relative body motion,

followed by absolute body motion, and apparent motion in each of the pixels in the video, suggesting that the prediction may originate at a higher level (i.e., how the ballet figure being performed will unfold, independent of viewpoint), and progress to become more concrete towards the time at which the actual sensory input needs to be compared to the prediction. Indeed, a posthoc exploratory ANOVA on the jackknifed peak latencies with factors model (3) and ROI (4, i.e., ROIs with significant dRSA curves in all 3 models), confirmed a main effect of model at $p = 3.2 \times 10^{-9}$ (Fig. 3a). While predictive representations were present in all regions of the AON, they were strongest in visual areas, which was further confirmed by a searchlight analysis in which dRSA was run across the whole cortical surface (Fig. 2b).

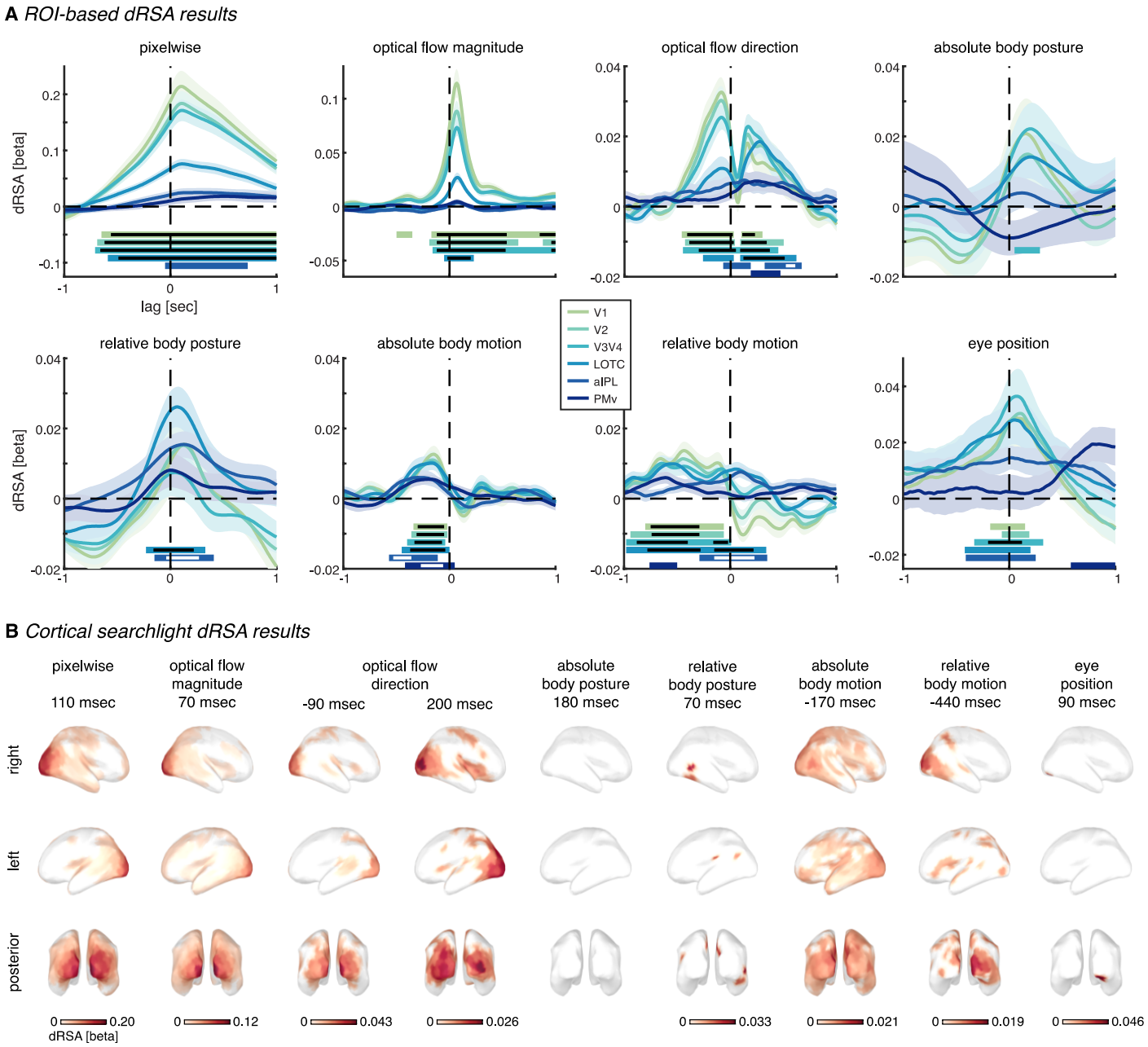


**Figure 2. dRSA results.** (a) ROI-based analysis. Dynamic RSA regression weights, illustrated as lag-plots after averaging the 2-dimensional dRSA matrix over time, where a positive lag indicates that the neural representation is lagging behind a model representation, while a negative lag indicates that the neural representation is preceding a model representation (i.e., a predictive neural representation). Different stimulus-feature models are plotted in separate subplots, while ROIs are plotted in different colors, with light-to-dark colors reflecting posterior-to-anterior ROIs. Lines and shaded areas indicate subject average and SEM, respectively. Horizontal bars indicate significant beta weights ($p < 0.01$ for single time samples), cluster-corrected for multiple comparisons ($p < 0.05$ for cluster-size), with colors matching the respective ROI line plots. Black or white horizontal bars inside colored bars indicate significance at a stricter statistical threshold (i.e., $p < 0.001$ for single samples and $p < 0.01$ for cluster-size). For full two-dimensional model-by-neural time plots see Fig. S5. (b) Searchlight on cortical surface. The dRSA analysis was run on all 400 parcels of the Schaefer atlas[24]. Peak timings were selected per model based on the ROI analysis illustrated in Figure 2. Only parcels with dRSA values surviving FDR correction are displayed ($p < 0.05$ for a one-sided t-test). For uncorrected source maps see Fig. S6.

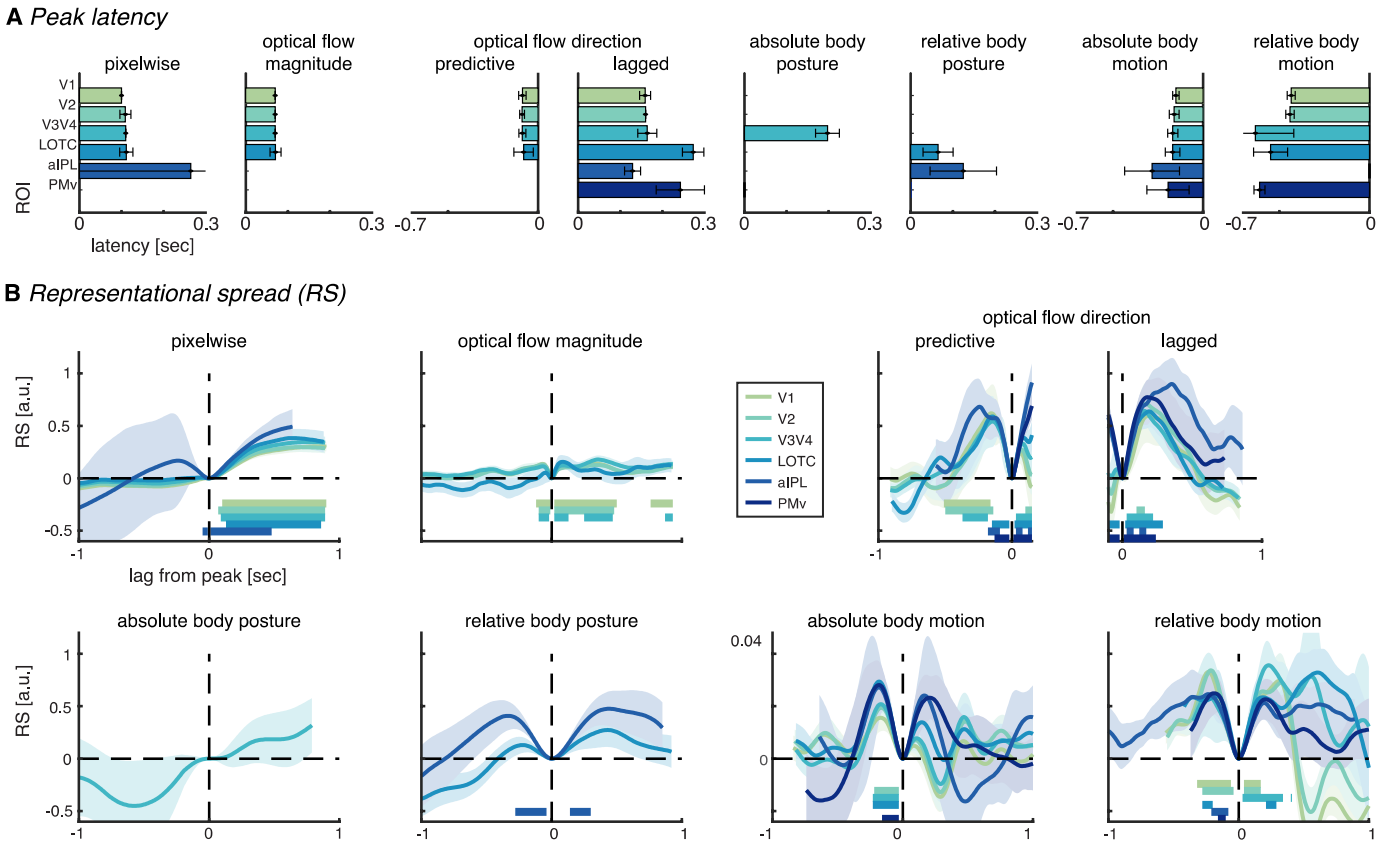


**Figure 3. Peak latencies and representational spread.** Only results are plotted for ROI-model combinations that showed a significant representation in the main dRSA analysis in figure 2. a) Peak Latencies of dRSA curves in figure 2. Bars and error bars indicate jackknifed subject average and SEM, respectively. b) Representational spread (RS), which indicates temporal information spread surrounding the dRSA peak, above and beyond what can be explained by model autocorrelation. The horizontal axis indicates temporal distance from the corresponding peak indicated in figure 2, with zero lag coinciding with the peak latency in figure 2. Positive lag indicates representational stability on the right (downward) slope in figure 2, while negative lag indicates stability on the left (upward) slope in figure 2. Lines and shaded areas indicate jackknifed subject average and SEM, respectively, Horizontal bars indicate significant RSI ($p < 0.001$ for single time samples), cluster-corrected for multiple comparisons ($p < 0.01$ for cluster-size), with colors matching the respective ROI line plots. In both (a) and (b) different stimulus-feature models are plotted in separate subplots, with light-to-dark colors reflecting posterior-to-anterior (or low-to-high level) ROIs. Note that as we observed both a predictive and lagged peak for optical flow direction (Fig. 2), we computed RSI separately for each peak.

## Neural representations are sustained

Interestingly, most lagged representations (i.e., pixelwise, optical flow magnitude and direction, and relative body posture) show a similar asymmetry in that the slope is less steep towards the lagged (right) side (Fig. 2a), which cannot be explained by the model autocorrelations, as those are symmetrical (Fig. S4a and b). In a posthoc exploratory analysis we directly compared the amount of information spread in the neural signal surrounding a peak, with the amount that can be expected based on model autocorrelation alone. Specifically, for each model we normalized both the observed and simulated dRSA curves by dividing each by its peak value, and aligned the curves, such that each had a peak latency of zero and peak value of 1. Last, we computed the representational spread (RS) by subtracting the simulated from the observed curve, thus quantifying the relative amount of information spread surrounding a dRSA peak, above and beyond what can be accounted for by the model autocorrelation (Fig. 3b). The RS indicated that stimulus information is not represented in the brain at a single time point, but rather is represented across a sustained temporal window[4], with more prominent temporal spreading towards the lagged (right) side for most lagged representations, more spreading towards the predictive (left) side for most predictive representations, and spreading in both directions for relative body motion.

## Low level static and motion stimulus representations in two separate pathways

Low-level stimulus information was most strongly represented in visual areas with a delay relative to the actual input, both for static pixelwise luminance information (100–110 msec) and for the magnitude of apparent motion in terms of optical flow vectors at each pixel (70–80 msec). The fact that the static pixelwise model showed a slower response, that was furthermore relatively sustained (see gentler slope on the right versus left side of the peak in figure 2a and 3b), while the low-level motion model showed a faster and more transient response, is in line with a separation of two processing pathways. That is, while the parvocellular pathway processes static information at high spatial resolution with a slow and sustained response pattern, the magnocellular pathway processes motion information with a faster and transient response pattern[25]. The difference in peak latencies was confirmed by a posthoc exploratory ANOVA with factors model (2) and ROI (4, i.e., ROIs with significant dRSA curves in both models), showing a main effect of model at $p = 9.2 \times 10^{-6}$ (Fig. 3a).

**Body posture and motion engage the AON**

We observed a progression from low to high complexity models being represented in posterior to anterior ROIs, respectively. Specifically, while the lagged representations of pixelwise and optical flow magnitude were most strongly represented in visual areas, the lagged representations of body posture and optical flow direction similarly engaged nodes of the AON. This is most evident for relative body posture, arguably a high-level representation closely related to a conceptual representation, which was represented most strongly in LOTC, corroborating LOTC's central role in action recognition[26,27]. The searchlight analysis (Fig. 2b) confirmed this spatial pattern, and further indicated that it was LOTC specifically in the right hemisphere that most strongly represented relative body posture.

## Discussion

We investigated the rich dynamics of predictive neural representations of naturalistic sensory input, by means of a novel dynamic extension to RSA that uses temporally variable stimulus models. We applied this new approach to source-reconstructed MEG data of healthy human subjects observing ballet dancing sequences. Dynamic RSA revealed novel insights into the dynamics of neural representations of low-level visual, body posture and motion features: First, most interestingly, body motion was represented in a temporal window that preceded the actual input, indicating that these neural representations predicted future motion. Furthermore, predictive representations showed a hierarchical pattern, such that higher-level stimulus features were predicted earlier in time, while lower-level features were predicted closer in time to the actual sensory input. Second, both low- and high-level information was represented after the actual visual input, with a progression from low to high complexity models being represented in lower visual areas to more anterior nodes of the AON, respectively. Third, we observed a dissociable pattern in the representations of static versus dynamic low-level visual information (i.e., pixelwise vs. optical flow magnitude), in that static visual information showed a slower and sustained response, whereas dynamic low-level information showed a faster and transient response, in line with a parvocellular and magnocellular pathway, respectively[25]. These latter two patterns agree well with the literature on visual processing, and demonstrate that dRSA can be used to characterize different processing pathways, e.g., by providing a window into the temporal dynamics of visual processing of naturalistic dynamic input: stimulus information is not immediately overwritten by new input, but instead is represented across a sustained temporal window. Taken together, this promising new approach opens the door for addressing important outstanding questions on how and when our brain represents and predicts the dynamics of our world.

Interestingly, while most obvious for the pixelwise model, all lagged representations we observed were more temporally sustained towards the lagging tail. Such sustained representations may form the basis for neural predictions. That is, predicting the future requires to integrate past information across an extended period in order to extrapolate this information[4], e.g., for predicting the future position of moving objects[29]. This might even be an emergent property, as velocity-tuned neurons in an unsupervised hierarchical neural network spontaneously start to extrapolate the position of a moving object along its trajectory[30]. This property likely generalizes across sensory domains, as our auditory system integrates past tone information to form predictions of future tones in tone sequences with naturalistic statistical regularities[14]. As such, the sustained nature of the lagged representations observed here might be necessary for creating and updating predictions based on constantly changing information in our dynamic world. An alternative function of sustained neural representations might be the integration of features to form coherent higher-level representations. That is, as different features have different processing times, an overlapping temporal window of representation is needed for their integration[4]. Future research should investigate how and which past information is integrated exactly to form predictions of naturalistic input, e.g., as done for simple tone sequences[14], or by investigating how long it takes for predictions to start up again after feature-specific perturbations.

We specifically observed predictive representations of the future motion direction of the ballet dancer, both as defined by the direction of optical flow vectors that capture the motion at each pixel, and the kinematic markers that capture a higher-level motion representation. While there is ample research on the prediction of future position of moving objects, the prediction of future motion direction is largely unexplored. Nevertheless, several studies did report prediction of motion direction; e.g., if an auditory cue reliably predicts the direction of a moving dot pattern, both perception and neural representation are biased towards the predicted direction[11]. Furthermore, responses in V1 and hMT/V5+ are reduced when the direction of a random-dot motion can be implicitly predicted by the surrounding context outside of the population receptive field (pRF)[31], and simulated neural populations automatically encode motion direction when optimized for prediction[32]. However, these studies use simple unnaturalistic stimuli of which the predictability is manipulated. It is more difficult to investigate the representational dynamics of motion direction in complex naturalistic stimuli, where predictions are formed implicitly. With our new approach, we show for the first time the prediction of future body motion. Interestingly, we observed a hierarchical pattern in motion prediction in terms of stimulus model complexity, where relative body motion was predicted earlier than absolute body motion, and the direction of optical flow vectors even later. While speculative, it might be that a prediction starts at a high conceptual level of how the observed action will unfold, which is best captured by the relative body motion model. This subsequently translates into a more concrete prediction, i.e., absolute body motion, and in turn the optical flow vector direction, to which the sensory input can be directly compared. Such hierarchical prediction is well in line with predictive processing theories[22,23], and might in fact be an intrinsic property of cortical circuitry[33]. Indeed, ample empirical evidence demonstrates how high level context information shapes low level predictive representations[1,34],

in vision[35], audition[18,19], and natural language processing[36,37]. Important here, hierarchical predictions are also evident in complex contexts such as social perception and action observation[38,39]. While these previous studies demonstrate the hierarchical nature of predictions, they do not look at the predictive representations themselves. Here we show the rich dynamics of hierarchical predictive representations at three levels of processing of naturalistic dynamic stimuli.

Interestingly, while ample previous studies found predictive processing of future object position[6,7,10], we did not observe a predictive representation of future body posture. However, we argue that the lagged representation of body posture observed here might not reflect pure bottom-up processing, but instead (largely) reflect prediction. One would expect a more complex representation to have a larger processing latency compared to lower-level representations. In contrast, here we observe a peak latency for relative body posture in LOTC that is earlier than the peak of the absolute body posture and pixelwise models, suggesting some form of prediction has shifted the peak of the relative body posture closer to real-time. In other words, while a negative peak latency in dRSA evidences prediction, a positive peak latency precludes inferring a pure bottom-up representation. The idea that the body posture representations might also (largely) reflect prediction is corroborated by several recent findings demonstrating the tenacity of predictive processing in the brain, such as the tendency for representations to be skewed towards future states[40], the fact that top-down connections seem essential to model human visual processing[41], and the importance of such connections for action prediction[12]. Future research utilizing dRSA can discern this by manipulating stimulus predictability. An explanation for the more obvious predictive representation for motion compared to position is that we might in fact predict future motion most strongly in naturalistic dynamic stimuli, perhaps because of our rich experience with constrained end points in biological motion (i.e., end points are preceded and succeeded by motion towards and away from the end point, respectively). This would make motion prediction in these highly predictable biological stimuli particularly reliable. Motion prediction may have been further reinforced by the task, which encouraged attention specifically towards the ballet dancer's motion, rather than position. Future research using dRSA can discern this by experimentally manipulating attention.

Finally, the dRSA approach represents a methodological advancement for studying when and how our brain represents and predicts the dynamics of our world. It uses naturalistic dynamic stimuli and dynamic stimulus models and reveals both the peak latency and stability of neural representations, and it does so at several hierarchical levels of processing by characterizing stimuli using different (computational) stimulus models to capture different stimulus features. In principle, this approach can be used to study the rich representational dynamics of naturalistic stimuli in any context (e.g., naturalistic soundscapes, music, language, motor planning/execution, social interaction), and in any organism or signal with high enough temporal resolution (e.g., M/EEG, EMG, EcoG, animal electrophysiology). Several recent studies investigated the dynamics (and predictive nature) of neural representations using related approaches. For instance, the temporal generalization of decoding approach captures if and how patterns of neural activity that code for stimulus information generalize between time points[42]. However, this approach does not distinguish between stimulus features (i.e., decoding analysis is blind as to which information is used to separate classes), nor can it capture dynamic features, as class definitions would change over time. Other approaches use RSA to study prediction or top-down processing, for instance by computing representational similarity between different ROIs at different time points to estimate the directionality of information flow[41], or by using RSA to distinguish words prior to their presentation during naturalistic reading[13]. All these approaches have in common that they use highly controlled static stimuli and stimulus models. The new dRSA approach is unique as it allows studying dynamic naturalistic input.

To conclude, we present a novel methodological framework that unveils at which point in time relative to stimulus presentation different stimulus features are represented in the brain, and reveals predictive neural representations of the future motion of an observed ballet dancer.

## Materials and Methods

### Subjects

Twenty-two healthy human volunteers (mean age, 30 ± 7 years; 13 females) participated in the experiment for monetary compensation. All subjects were right-handed, had normal or correct-to-normal vision, and were naive with respect to the purpose of the study. All experimental procedures were performed in accordance with the Declaration of Helsinki and were positively reviewed by the Ethical Committee of the University of Trento. Written informed consent was obtained. The entire session including preparation lasted ~3h, of which 1.5h were spent in the MEG scanner.

### Stimuli

The video stimulus set consisted of 14 unique ballet sequences, each consisting of four smoothly connected ballet figures selected from a pool of 5 unique figures (i.e., passé, pirouette, arabesque, bow, and jump; see Table S1 for the structure of each unique sequence). All figures were presented from multiple viewpoints, to allow for studying absolute body posture and motion separately from relative body posture and motion (i.e., orientation and position invariant). The pirouette could start and end at different angles and was always performed counter-clockwise. The rational for this particular stimulus design was that 1) at each point in time there should be enough stimuli similar and enough stimuli dissimilar (i.e., enough variance within a model RDM at a single time point), and that 2) the model RDMs should be sufficiently dynamic (i.e., minimal temporal autocorrelation) for testing lag differences between model and neural RDMs. We recorded the ballet

dancing videos with a Canon PowerShot G9 X Mark II at a frame rate of 50 Hz (i.e., 250 frames per 5 second video). Video frames were cropped to 400x376 [height x width] pixels. Additionally, we recorded 3-dimensional positions of 13 passive (infrared reflecting) markers placed on the ankles, knees, hips, shoulders, elbows, wrists and head of the ballet dancer. Marker positions were recorded at 100 Hz using a Qualisys motion capture system (infrared-based marker capture). Markers were small silver balls, and nearly invisible in the video stimuli presented to the subjects. See https://github.com/Ingmar-de-Vries/DynamicPredictions for all 14 unique videos and temporally aligned 3D kinematic data.

**Experimental design**

The experiment was created using the Psychophysics Toolbox (version 3) in MATLAB (version 2012b; MathWorks). Trials consisted of 5 second ballet dancing videos (Fig. S1a), and were separated by blank screens with a white fixation cross presented for 1.8 – 2.2 seconds (uniform distribution). The fixation cross was also presented throughout the video, and subjects were instructed to keep fixation while covertly attending to the ballet dancer. A single run in the MEG scanner lasted ~12 minutes and consisted of 90 randomly ordered sequences with the only constraint that the exact same sequence was never directly repeated (i.e., there was always at least one different sequence in between).

To ensure attention to the ballet sequences, 20 trials in each run of 90 trials were two types of catch trials (Fig. S1b). In type 1 (15 trials) the sequence was interrupted at a random time between 0.5 – 5 seconds after onset, and subjects were asked with a yes or no question (2 response buttons) about the particular motion currently being performed (e.g., *arms moving up?* or *pirouette?*), and had to respond with a button press within 3 seconds. In type 2 (5 trials) the fixation cross changed from white to light purple (RGB = [117 112 179]) for 200 msec at a random time between 0.4 and 4.8 seconds after onset, in response to which subjects had to press a button. The color was chosen such that the change was only visible when fixating, but not from the periphery, thus stimulating subjects' fixation. Subjects received immediate feedback on each catch trial. In 5 out of the 15 type-1 catch trials, the question appeared only at the offset of the video (i.e., at 5 seconds), and these trials were therefore included in the analyses, thus leaving 75 to-be-analyzed trials per run. The remaining catch trials were excluded from all further analyses. To ensure continuous attention to the task, two subsequent catch trials were never separated by more than 8 normal trials. Subjects performed one practice run outside of the MEG scanner, in which 30% of trials were catch trials to ensure enough practice with the task. During MEG, subjects performed the behavioral task well, indicating that they attended to the ballet dancer and generally kept fixation (mean/min-max: 80/63-93% correct and 95/80-100% detection for the dancer and fixation catch trials, respectively; see Fig. S1c). Subjects performed 6.8 runs on average depending on attentional fatigue, which resulted in an average of 510 trials per subject for further analyses, with an average of 36.4 repetitions of each of the 14 unique sequences.

**Stimulus models**

We selected a total of 9 models described in detail below that each capture features of the ballet sequences at different hierarchical levels of complexity (see Figure S2 for an overview), plus a tenth subject-specific model capturing the eye position. Models based on video data (pixelwise and optical flow vectors) had a sampling frequency of 50 Hz, and were therefore first interpolated to 100 Hz to match the kinematic data, using shape-preserving piecewise cubic interpolation.

Pixelwise (Fig. S2b): For each video frame, the RGB values at the 400x376 pixels were converted to grayscale (i.e., luminance values), smoothed (i.e., spatially low-pass filtered) using a Gaussian kernel with sigma = 5 (i.e., roughly 11.75 pixels full width at half max[20]), as implemented in MATLAB's `imgaussfilt.m` function, and vectorized (i.e., 150400 feature vector per frame). For each frame separately, these vectors were then correlated between two different videos, to generate a single entry in the model RDM, where dissimilarity was defined as 1-Pearson correlation. This operation was repeated for each of the possible stimulus pairs to generate a 14x14 RDM at a single frame, and subsequently repeated over the 500 frames to generate 500 model RDMs (see Fig. 1e for an example).

Optical flow vectors (Fig. S2c): We converted videos to grayscale and used MATLAB's `estimateFlow.m` implementation of the Farnebäck algorithm[43] to compute optical flow vectors, which capture the apparent motion between subsequent frames at each pixel. These optical flow vectors were spatially smoothed using a Gaussian kernel[44], with the same settings as for the pixelwise grayscale model, and subsequently divided into magnitude and 2D direction, and vectorized (i.e., 150400 and 300800 feature vectors for magnitude and direction, respectively). Model RDMs were computed as described for the pixelwise model.

Absolute body posture (Fig S2d): We defined absolute body posture as the 3D positions of the 13 kinematic markers (i.e., 39 feature vector), relative to a coordinate system with origin placed on the floor in the center of the video frame. Model RDMs were computed as described for the pixelwise model.

Relative body posture (Fig. S2d): For the relative body posture model, before computing the dissimilarity between two postures in two different stimuli, we used a modified version of MATLAB's `procrustes.m` function to align the 3D marker structure of one stimulus with the marker structure of the other stimulus, while keeping its internal structure intact, by using translation, and rotation along the vertical axis only, to minimize the sum of squared differences between the two 3D marker structures. For instance, if in two to-be-compared videos the exact same ballet figure is performed but in opposite direction, this would result in high dissimilarity in terms of absolute body posture, but high similarity in terms of relative body posture. Since the direction of alignment matters for dissimilarity (i.e., dissimilarity(frame1,frame2aligned) ≠ dissimilarity(frame1aligned,frame2)), we computed dissimilarities after alignment in both directions, and subsequently averaged over them.

Absolute body motion (Fig. S2d): Absolute body motion was operationalized as the first derivative of absolute body posture, i.e., the difference in the 3D marker positions between two subsequent frames: $v_m(i) = x_m(i) - x_m(i-1)$, where $v$ is the motion of marker $m$ in frame $i$, and $x$ is the marker position. Body motion vectors therefore had a magnitude and a 3D direction. Motion time series were shifted and interpolated such that the motion calculated in the formula above for frame $i$ was actually placed in between frame $i$ and frame $i-1$. Model RDMs were computed as described for the pixelwise model.

Relative body motion (Fig. S2d): Relative body motion was computed the same way as absolute body motion, but after aligning one body posture to the other as described above for relative body posture. Model RDMs were computed as described for the pixelwise model.

Eye position: Since eye movements are a common problematic covariate in neuroimaging studies using multivariate approaches[28], even if like here a fixation task is used (e.g., micro-saccades remain), we added individual-subject RDMs based on eye-position as a control to ensure that the dRSA results for our models of interest were not due to eye movements. The raw eye tracker signals were first downsampled to 100 Hz, after which trials in which the eye tracker lost the eye for more than 10% of the samples were removed, and the remaining missing samples (e.g., due to blinks) were interpolated. Finally, signals were averaged over all remaining repetitions of the same unique stimulus. Eye position models were subject-specific. Model RDMs were computed as described for the pixelwise model, except that dissimilarity between eye positions was defined as the Euclidean distance between the four eye position values (i.e., x-, and y-position for left and right eye), as four values is too low for a reliable correlation estimate. Note that while there was a representation of eye position present in the neural signal (Fig. 2), this was successfully regressed out when testing each of the other models (as confirmed by the simulations; Fig. S4b). Therefore, eye movements could not explain any of the findings for the other models.

Absolute and relative body acceleration were operationalized as the second derivative of marker position (i.e., first derivative of marker motion). These models only explained a minimal amount of variance in the neural data, and are therefore not illustrated in main results. However, because these models did correlate with some of the tested models (Fig. S4a), we regressed them out when testing each of the other models (see '*Dynamic representational similarity analysis'* for details on regression).

**MEG data collection and preprocessing**

We obtained whole-head MEG recordings at a sampling rate of 1000 Hz using a 306-channel (204 first-order planar gradiometers, 102 magnetometers) VectorView MEG system (Neuromag, Elekta) in a two-layer magnetically shielded room (AK3B, Vacuum Schmelze). A low-pass antialiasing filter at 330Hz and a DC offset correction were applied online. Importantly, no online or offline high-pass filter was applied, since high-pass filtering can temporally spread multivariate information[45]. Before the MEG recording, we digitized the individual head shape with an electromagnetic position and orientation monitoring system (FASTRAK, Polhemus) using the positions of three anatomic landmarks (nasion and left and right preauricular points), five head position indicator coils, and 300+ additional points evenly spread on the subject's head. Landmarks and head-position induction coils were digitized twice to ensure a localization error of < 1 mm. To co-register the head position in the MEG helmet with anatomic scans for source reconstruction, we acquired the head positions at the start of each run by passing small currents through the coils. Binocular eye-movements were recorded at 1000 Hz using an SR-Research Eyelink Plus eye-tracker. In addition, horizontal and vertical EOGs were recorded from 2 electrodes located 1 cm lateral to the external canthi and 2 electrodes located 2 cm above and below the right eye, respectively. Stimuli were presented in full color using a VPixx PROPixx projector. All hardware was connected to a DataPixx input/output hub to deliver visual stimuli and collect MEG data, eye-tracker data and button presses in a critical real-time manner (VPixx Technologies), and stimulus-onset triggers were directly stored together with the MEG data. In addition to these triggers, we also placed a photodiode in the top right of the screen (i.e., outside of the video stimulus frame), and changed a square on the screen underneath the photodiode from black to white at the onset of each single video frame, as to realign MEG signals to these photodiode signals offline.

MEG data were preprocessed and further analyzed offline using a combination of the Brainstorm[46], Fieldtrip[47] and CoSMoMVPA[48] toolboxes in MATLAB, as well as custom-written MATLAB scripts (any scripts that are not standard toolbox or MATLAB implementations are shared at https://github.com/Ingmar-de-Vries/DynamicPredictions). Continuous data from each run were visually inspected for noisy sensors and system-related artifacts (e.g., SQUID jumps), and a maximum of 12 noisy sensors were removed and interpolated. Next, we applied the Neuromag MaxFilter implementation of Signal Source Separation (SSS[49]) for removing external noise to each individual run and spatially aligned the head position inside the helmet across runs. After these initial steps, data were loaded into Brainstorm, where consecutively spatial co-registration between anatomical MRI and head position in the MEG helmet was refined using the 300+ additionally registered points on the head (see *'Source reconstruction'*), data were filtered for line noise using a band-stop filter at 50 and 100 Hz (0.5 Hz bandwidth), data were down-sampled to 500 Hz, and independent component analysis (ICA) was applied separately for magneto- and gradiometers in order to detect and remove an average of 1.2 blink and eye movement components and 0.6 cardiac components. Blink and eye movement components were confirmed by comparing their time series with the EOG channels. ICA was applied to a temporary version of the data that was first down-sampled (250 Hz) and band-pass filtered (0.1 – 100 Hz) to reduce computation time and improve ICA, respectively. The ICA weight matrix was subsequently applied to the original data. Note that the maximum number of ICA components to be extracted was determined by the residual

degrees of freedom after SSS rank reduction during MaxFiltering. Continuous data were then segmented in epochs from -1.5 to 6.5 seconds locked to the onset of the video stimulus, according to the triggers stored with the MEG data, and single epochs were DC offset corrected using a baseline window of -500 to 0 msec relative to stimulus onset. Each epoch was visually inspected, and those containing artifacts (e.g., muscle activity, SQUID jumps, etc.) were discarded from further analyses, which resulted in a rejection of 7.2% of all trials. After these steps, data were exported from Brainstorm into the Fieldtrip data format, epochs were realigned to the photodiode onset for higher temporal precision, data were temporally smoothed using a 20 msec boxcar kernel as implemented in Fieldtrip's `ft_preproc_smooth.m`, downsampled to 100 Hz, and finally averaged over all ~36 repetitions of the same ballet video stimulus. Note that we opted for transforming MEG signals to source space (see next section), but if one would opt to stay in sensor space instead, we recommend applying multivariate noise normalization (MNN[50]), which we confirmed improves dRSA at sensor level (data not shown).

**Source reconstruction and ROI selection**

For the construction of 3D forward models, we first acquired previously recorded anatomical 3D images if available (13 subjects), and used the ICBM152 standard brain for the remaining 9 subjects. Individual anatomical images were obtained using a Siemens Prisma 3T, actively shielded, whole body 3 Tesla magnet. A few older individual anatomical scans already present at the facility were collected with the former scanner, i.e., 4T MRI scanner (Bruker Biospin). Fiducial points were marked automatically (MNI) and inspected, after which the anatomical scans were 3D reconstructed using CAT12, a computational anatomy toolbox[51]. For the 9 subjects without individual anatomy, the standard anatomy was warped to the subject's head volume as estimated from the +300 digitized head points. Next, head models were computed by coregistering the subjects head shape with the reconstructed MRI brain volumes using Brainstorm's default settings and using overlapping spheres, after which we performed source reconstruction using minimum-norm estimates (MNE) for a source space of 15000 vertices[52], as implemented in Brainstorm. We used both gradio- and magnetometers. For MNE, first noise covariance matrices were computed after single-trial DC offset correction, and using the time window from -1000 to 0 msec relative to stimulus onset. Second, the data covariance matrices were computed using the -1000 to 0 msec interval for baseline correction, and the 0 to 5000 msec interval (i.e., stimulus presentation) as data interval. Last, we computed non-normalized current density maps as a measure of source activity, with source direction constrained to be normal to the cortex (i.e., a single source activity value at each vertex).

The 15000 vertices were parcellated into 360 parcels according to the cortical atlas from the Human Connectome Project (HCP[53]), which provides the most precise parcellation based on structural and functional organization of the human cortex to date. We a priori selected 6 regions of interest (ROIs) that consisted either of a single parcel from the atlas (in case of V1 and V2), or a combination of several parcels (in case of V3+V4, the lateral occipitotemporal cortex (LOTC), the anterior inferior parietal lobe (aIPL), and the ventral premotor cortex (PMv); see Fig. S3 for details). Besides the purely visual areas, we selected the LOTC, aIPL and PMv because they form the action observation network (AON[54]), and were expected to be important for processing and predicting these specific stimuli[39,55]. This selection resulted in comparably sized ROIs with on average 484 vertices (min–max = 403–565), of which the vertex signals were used as features for computing the neural RDM for a given ROI. The neural RDM was computed as implemented in cosmoMVPA[48], on the centred data, using 1-Pearson correlation as dissimilarity measure.

**Dynamic representational similarity analysis**

As a first step, neural and model representational dissimilarity matrices (RDMs) were computed for each time point as described above (see also Fig. 1), resulting in neural and model RDMs of 91 features (14x14 stimuli gives 91 unique values) by 500 time points. To improve the signal-to-noise ratio, neural and model RDMs were temporally smoothed using a 20 msec boxcar kernel as implemented in Fieldtrip's `ft_preproc_smooth.m`. To compute the dRSA, the similarity between neural and model RDMs was estimated at each neural-by-model time point, thus resulting in a 2-dimensional dRSA matrix (Fig. 1f; lower panel). This 2D dRSA matrix was further averaged over time for each neural-to-model time lag (i.e., time-locked to the on-diagonal) within the range of -1 to 1 second, thus resulting in a lag-plot (Fig. 1; upper panel), in which a peak on the right of the vertical zero-lag line indicates that the neural RDM is most similar to a model RDM earlier in time (i.e., it is lagging behind the stimulus information), whereas a peak on the left of the vertical zero-lag line indicates that the neural RDM is predicting a future model RDM. One problem with using a simple correlation as similarity measure for RSA is that different stimulus models might share variance (e.g., absolute and relative body posture will inevitably share some variance), making it impossible to unambiguously assign an RSA result to one specific model. Additionally, specific to dynamic RSA, model RDM X at $t_1$ might share variance with model RDM Y at $t_2$ (e.g., body motion will correlate with body posture at an earlier and later time point), and model RDM X might additionally share variance with itself at distant time points due to temporal autocorrelation. These problems are illustrated for simulated data in Figure S4a (also see *"Simulations"*).

Therefore, in order to extract variance from the neural data uniquely explained by a specific stimulus model, while disregarding variance better explained by other models, we used linear least-squares regression to estimate similarity. Here, the neural RDM at a certain time point $t_y$ was the response variable Y (size 91x1), our model RDM of interest at a certain time point $t_x$ was one predictor variable $X_{test}$ (size 91x1), and all to-be-regressed out RDMs were other predictor variables $X_{regressout}$ (size 91xN), with our measure of similarity being the regression weight for $X_{test}$. As to-be-regressed out RDMs we

selected the other 9 model RDMs at each time point (i.e., size 91x$N_{time}$ per model), plus the model RDM of interest itself at distant time points (i.e., at a minimum distance from $t_x$ at which the model shared less than 10% variance with itself at $t_x$; determined by the simulations illustrated in Fig. S4a). This would result in >1800 predictor variables in $X_{regressout}$ (i.e., size 91x1800). However, the amount of predictor variables in linear least-squares regression is limited by the number of features (also called observations; here a maximum theoretical value of 91, but one should stay below this maximum to prevent overfitting). We therefore used principal component analysis (PCA) across two steps to reduce the dimensionality of $X_{regressout}$. First, we ran PCA separately per to-be-regressed out model (size 91x$N_{time}$), and selected only those principal components capturing at least 0.1% of total variance from the $N_{time}$ input variables. Next, the new $X_{regressout}$, now consisting of principal components, was combined with $X_{test}$, after which a second PCA was run over this combination. From this second PCA, only the first 75 components (ordered by total explained variance) were selected (limited by the maximum of 91 predictors), and these components were used as predictor variables in the linear least-squares regression. Subsequently, the principal component regression weights were projected back onto the original predictors using the PCA loadings, in order to specifically extract the regression weight for $X_{test}$. This approach, termed principal component regression (PCR), has the advantage that it strongly reduces the number of predictors while maintaining most of their total variance, and the advantage that the PCA components entered into the regression are decorrelated, hence diminishing the risk of multicollinearity. Before each PCA, individual model RDMs and the neural RDM were rescaled (to min-max: 0-2) at once for all time points, to equalize scales between RDMs, while keeping the temporal changes in scale intact that are inherent to a certain model. Additionally, the RDM was centred per individual time point before each PCA and regression. Importantly, PCR was effective at extracting a model representation of interest while regressing out other model representations with which it might share variance, as confirmed by simulations (see '*Simulations*' and Fig. S4b, c and d).

Our dRSA pipeline involved one additional important step to increase sensitivity. If computing dRSA in the thus far described manner, the 2D dRSA matrix structure is very inhomogeneous. This is because the structure of a dynamic model RDM changes depending on dynamic feature trajectories inherent to specific stimuli (e.g., event boundaries at different timescales[56], such as motion onsets and offsets, video onset, etc.). This causes a heterogeneous pattern across video time in terms of variance in the model RDM and in terms of stimulus predictability for each specific stimulus. To minimize heterogeneity of RDM variance and predictability, we used a temporal resampling method (see box around Fig. 1d, e and f). Specifically, across 1000 iterations we randomly selected a different 3-second segment for each of the 14 stimuli (Fig. 1d), computed the RDMs (Fig. 1e) and dRSA (Fig. 1f) as described above on all 1000 iterations, and averaged dRSA matrices across iterations. These averaged dRSA values were used for subsequent statistical analysis (see '*Statistical analysis*').

In order to successfully apply dRSA, it is worth considering a few criteria that the model RDMs ideally fulfil. Specifically, they need to be sufficiently large (i.e., have enough unique stimuli) to allow for enough regressors to be included in the PCR step, sufficiently large such that results are generalizable rather than specific to the stimulus set (which we solved here with the resampling procedure), and dynamic enough that the structure of the RDM at different time points is distinguishable (i.e., stimulus pairs need to be highly similar at some times while dissimilar at other times).

**Statistical analysis**

Before group-level statistics (and plotting), we Fisher Z-transformed all beta weights to get normally distributed values[57]. Because statistical testing of the dRSA lag-plots involves many comparisons (at each lag time-point), we performed group-level nonparametric permutation testing with cluster-based correction for multiple comparisons[58], as implemented in Fieldtrip's `ft_timelockstatistics.m` function, with the method parameter set to "montecarlo". We used t-values from a one-sided t-test as test statistic (i.e., to test whether beta weights are significantly larger than zero), we used the sum over the t-values belonging to a cluster of significant time points as measure of cluster size, and we ran the permutation 25000 times for each analysis. We selected a lenient and strict threshold for significance, with a single-sample threshold at $p < 0.01$ or $p < 0.001$, combined with a cluster-size threshold at $p < 0.05$ or $p < 0.01$, respectively. Only significant clusters larger than what can be expected by chance survive this procedure.

**Simulations**

The goal of the simulations to find a dRSA pipeline that was effective in testing only a single model of interest while regressing out other models with which the model of interest might share variance, and while regressing out the model of interest itself at distant (future or past) time points (i.e., to reduce effects of temporal autocorrelation). Principal component regression (see '*Dynamic representational similarity analysis*') appeared effective to achieve this goal. Importantly, we first fine-tuned the dRSA pipeline described above on the simulations, after which we applied the exact same pipeline to the real neural data.

Each simulated neural RDM consisted of a single model RDM at zero lag. We first ran dRSA using a simple correlation for each tested model, i.e., the correlation coefficient was used as similarity measure (Fig. S4a). Note the nonzero off-midline correlations, which confirm a complex relationship between and within models across time. This complex structure of shared variance limits unambiguous inference if simple correlation is used for dRSA. Subsequently, the exact dRSA pipeline as described above using PCR was applied on each of the simulated neural RDMs (Fig. S4b), which indicated that PCR is

effective in extracting only the model of interest from the simulated neural RDM, while largely regressing out the other models. Note that not all models are completely regressed out, which is most obvious when motion models are tested on simulated data in which acceleration models are implanted. However, it is important to note that shared variance with another model is only problematic if that other model itself explains variance in the neural data. In other words, if model X explains variance in the neural data, and model Y does not, the observed results for model X cannot be explained by it sharing variance with model Y. Importantly, any of the remaining variance of a to-be-regressed out model Y that we observed while testing another model X (e.g., acceleration when testing motion; Fig. S4b), could not explain the main results for model X (e.g., motion; Fig. 2), as model Y itself did not explain enough of the neural data (here acceleration is not shown since this model only explained a very small portion of the neural data, but this logic works for all other models as well).

We performed two additional simulations to further test the effectiveness of PCR. First, we combined the first 3 models (pixelwise, optical flow magnitude and optical flow direction) with a random dynamic neural RDM (weight ratio 1:1:1:3) to create a single simulated neural RDM (Fig. S4c). The random neural RDM consisted of $91xN_{time}$ random values drawn from a uniform distribution, such that there was no temporal autocorrelation. As in the main dRSA pipeline, we tested each of the models separately, while regressing out the other 9 using PCR. The rational for selecting the first 3 models was that the first two models (pixelwise and optical flow magnitude) have a stronger representation in the real neural signal (Fig. 2), whereas the third model (optical flow direction) shows a predictive representation. Additionally, as all three are based on the video data, they might be more likely to share variance. We wanted to ensure that the predictive representation in the real neural signal for optical flow direction could not be caused by the presence of a representation for the first two models in the neural signal. Importantly, this analysis confirmed that PCR is effective in extracting a single model from the combined simulation, while regressing out the other two. Second, we ran dRSA on optical flow direction as simulated and tested model, but with different levels of added random noise (1:0, 1:1, 1:5, and 1:10) to create the simulated neural RDM (Fig. S4d). This analysis showed that simply adding more noise (i.e., variance unrelated to the model RDM of interest), can explain the values typically observed in the real neural data (Fig. 2).

**Peak latency**

Peak latencies were defined as the timing of the highest peak in the dRSA curves (Fig. 2a), separately for the negative (-1 to 0 sec) and positive (0 to -1 sec) intervals for the optical flow direction model (as we observed two distinct peaks), or at once for the whole -1 to 1 sec time interval for all other models. Given that peak latencies of the dRSA curves are difficult to estimate on a single-subject level (i.e., due to noise causing multiple peaks, similar to ERP/ERF analyses), we used a jackknife procedure[59]. Specifically, across 22 iterations we excluded one subject from the dataset, averaged dRSA curves over the remaining 21 subjects, and estimated the peak latency for this average dRSA curve, until each subject was excluded once. A problem with jackknifing is that because each jackknifed value now represents the average over n-1 subjects, the standard deviation across these values is artificially small. In order to use these jackknifed values in standard parametrical tests, we retrieved single-subject latency estimates by subtracting each individual jackknifed value multiplied by n-1, from the grandaverage jackknife value multiplied by n[60]. Essentially, rather than representing true individual subject dRSA peak latencies, these individual jackknifed values are a measure of how much a given subject's dRSA curve influenced the subject-average peak latency, while their average are a good estimate of the subject-average peak latency (Fig. 3a). Based on visual inspection of the main dRSA curves (Fig. 2a), we performed two sets of posthoc analyses on these jackknifed peak latencies. First, to test for a difference between pixelwise and optical flow magnitude, we performed two ANOVAs with factors model (2) and ROI (6 or 4), where we included either all 6 ROIs, or only the first 4 ROIs for which dRSA values were significant in the main analysis (Fig. 2a), as one could argue that peak latencies are unreliable if the dRSA curve itself is not significant. Second, we performed two similar ANOVAs with factors model (3) and ROI (6 or 4), to test for a difference in predictive peak latency for optical flow direction, absolute body motion, and relative body motion.

**Representational Spread (RS)**

The pyramidal shape of dRSA curves (Fig. 2a) could suggest a temporally extended (i.e., sustained) neural representation that is not completely overwritten by new visual input from subsequent video frames. However, this shape is partly explained by model autocorrelation (Fig. S4a), which also affects the shape of the dRSA curve resulting from the PCR method (Fig. S4b). We quantified the relative amount of information spread surrounding a dRSA peak, above and beyond what can be accounted for by the model autocorrelation, by means of a representational spread index (hereafter RS; Fig. 3b). Specifically, we first normalized for each model both the observed and the simulated dRSA curves by dividing each by its maximum (i.e., peak) value. Next, we aligned the curves such that they both had a peak latency of zero and peak value of 1. Last, we subtracted the simulated curve from the observed curve to retrieve the RS. As the RS relies on an accurate estimate of dRSA peak, we followed the exact same jackknife procedure described above for peak latencies. For this analysis, the simulated dRSA curve was computed using the PCR pipeline as for Fig. S4b, but with no added noise in the simulation, thus capturing the pure effect of the model autocorrelation in terms of temporal information spread in the dRSA curve.

**Searchlight on cortical surface**

To explore the spatial dRSA pattern across the cortical surface, we performed a whole brain analysis. For this analysis, we parcellated the 15000 sources in 400 parcels according to the Schaefer atlas[24], for which the parcels are more similar in

number of vertices compared to the HCP atlas (i.e., average ± std for HCP and Schaefer400 are 40 ± 33 and 36 ± 17, respectively). For each parcel, we computed dRSA as described above. To reduce computation time, we only computed dRSA at the neural-to-model time lag at which we observed a peak in the main ROI-based results (Fig. 2a), resulting in 9 separate searchlight analyses (i.e., 2 peaks for optical flow direction, and 1 peak for all other models; see Fig. 2b). For creating the cortical source maps, all vertices belonging to a certain parcel were given the same dRSA beta weight, hence giving each of 15000 sources a dRSA weight. These values were subsequently loaded into the Brainstorm GUI database and spatially smoothed using a Gaussian kernel with a full width at half maximum of 2 mm. We tested whether the dRSA beta weight at each of the 400 parcels was significantly larger than zero using one-sided t-tests, and used FDR correction at $p < 0.05$ to correct for multiple comparisons across the 400 parcels. Figure 2b illustrates the FDR corrected source maps, with parcels not surviving FDR correction colored in white, while Figure S6 illustrates the uncorrected maps.

## References


1. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).
2. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
3. Schubotz, R. I. Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* **11**, 211–218 (2007).
4. Hogendoorn, H. Perception in real-time: predicting the present, reconstructing the past. *Trends Cogn. Sci.* **26**, 128–141 (2022).
5. Bubic, A., Yves von Cramon, D. & Schubotz, R. I. Prediction, cognition and the brain. *Front. Hum. Neurosci.* **4**, 1–15 (2010).
6. Ekman, M., Kok, P. & De Lange, F. P. Time-compressed preplay of anticipated events in human primary visual cortex. *Nat. Commun.* **8**, 1–9 (2017).
7. Blom, T., Feuerriegel, D., Johnson, P., Bode, S. & Hogendoorn, H. Predictions drive neural representations of visual events ahead of incoming sensory information. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7510–7515 (2020).
8. Kok, P., Failing, M. F. & de Lange, F. P. Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *J. Cogn. Neurosci.* **26**, 1546–1554 (2014).
9. Kok, P., Jehee, J. F. M. & de Lange, F. P. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron* **75**, 265–270 (2012).
10. Hogendoorn, H. & Burkitt, A. N. Predictive coding of visual object position ahead of moving objects revealed by time-resolved EEG decoding. *Neuroimage* **171**, 55–61 (2018).
11. Aitken, F., Turner, G. & Kok, P. Prior Expectations of Motion Direction Modulate Early Sensory Processing. *J. Neurosci.* **40**, 6389–6397 (2020).
12. Cerliani, L. *et al.* Predictive coding during action observation – A depth-resolved intersubject functional correlation study at 7T. *Cortex* **148**, 121–138 (2022).
13. Wang, L. *et al.* Neural evidence for the prediction of animacy features during language comprehension: Evidence from MEG and EEG representational similarity analysis. *J. Neurosci.* **40**, 3278–3291 (2020).
14. Baumgarten, T. J. *et al.* Neural integration underlying naturalistic prediction flexibly adapts to varying sensory input rate. *Nat. Commun.* **12**, 1–14 (2021).
15. Schwiedrzik, C. M. & Freiwald, W. A. High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. *Neuron* **96**, 89-97.e4 (2017).
16. Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J. & Frith, C. The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* **7**, 413–422 (2012).
17. Urgen, B. A. & Saygin, A. P. Predictive processing account of action perception: Evidence from effective connectivity in the action observation network. *Cortex* **128**, 132–142 (2020).
18. Chao, Z. C., Takaura, K., Wang, L., Fujii, N. & Dehaene, S. Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. *Neuron* **100**, 1252-1266.e3 (2018).
19. Wacongne, C. *et al.* Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20754–20759 (2011).
20. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 1–28 (2008).
21. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455–462 (2014).
22. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
23. Rao, R. P. N. & Ballard, D. H. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
24. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
25. Merigan, W. H. & Maunsell, J. H. R. How Parallel are the Primate Visual Pathways? *Annu. Rev. Neurosci.* **16**, 369–402 (1993).

26. Wurm, M. F. & Caramazza, A. Two 'what' pathways for action and object recognition. *Trends Cogn. Sci.* **26**, 103–116 (2022).
27. Lingnau, A. & Downing, P. E. The lateral occipitotemporal cortex in action. *Trends Cogn. Sci.* **19**, 268–277 (2015).
28. Quax, S. C., Dijkstra, N., van Staveren, M. J., Bosch, S. E. & van Gerven, M. A. J. Eye movements explain decodability during perception and cued attention in MEG. *Neuroimage* **195**, 444–453 (2019).
29. Hogendoorn, H. Motion extrapolation in visual processing: Lessons from 25 years of flash-lag debate. *J. Neurosci.* **40**, 5698–5705 (2020).
30. Burkitt, A. N. & Hogendoorn, H. Predictive visual motion extrapolation emerges spontaneously and without supervision at each layer of a hierarchical neural network with spike-timing-dependent plasticity. *J. Neurosci.* **41**, 4428–4438 (2021).
31. Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W. & Muckli, L. Stimulus predictability reduces responses in primary visual cortex. *J. Neurosci.* **30**, 2960–2966 (2010).
32. Chalk, M., Marre, O. & Tkačik, G. Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 186–191 (2018).
33. Bastos, A. M. *et al.* Canonical Microcircuits for Predictive Coding. *Neuron* **76**, 695–711 (2012).
34. Gayet, S. & Peelen, M. V. Preparatory attention incorporates contextual expectations. *Curr. Biol.* **32**, 687-692.e6 (2022).
35. Kok, P. & De Lange, F. P. Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Curr. Biol.* **24**, 1531–1535 (2014).
36. Wang, L. *et al.* Dynamic predictive coding across the left fronto-temporal language hierarchy: Evidence from MEG, EEG and fMRI. *bioRxiv* 1–55 (2021).
37. Sohoglu, E., Peelle, J. E., Carlyon, R. P. & Davis, M. H. Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* **32**, 8443–8453 (2012).
38. Bach, P. & Schenke, K. C. Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Soc. Personal. Psychol. Compass* **11**, 1–17 (2017).
39. Kilner, J. M., Friston, K. J. & Frith, C. D. Predictive coding: An account of the mirror neuron system. *Cogn. Process.* **8**, 159–166 (2007).
40. Ekman, M., Kusch, S. & de Lange, F. P. Successor-like representation guides the prediction of future events in human visual cortex and hippocampus. *bioRxiv* (2022). doi:https://doi.org/10.1101/2022.03.23.485480
41. Kietzmann, T. C. *et al.* Recurrence is required to capture the representational dynamics of the human visual system. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21854–21863 (2019).
42. King, J. R. & Dehaene, S. Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn. Sci.* **18**, 203–210 (2014).
43. Farnebäck, G. Two-Frame Motion Estimation Based on Polynomial Expansion BT - Image Analysis. in (eds. Bigun, J. & Gustavsson, T.) 363–370 (Springer Berlin Heidelberg, 2003).
44. Efros, Berg, Mori & Malik. Recognizing action at a distance. in *Proceedings Ninth IEEE International Conference on Computer Vision* 726–733 vol.2 (2003). doi:10.1109/ICCV.2003.1238420
45. van Driel, J., Olivers, C. N. L. & Fahrenfort, J. J. High-pass filtering artifacts in multivariate classification of neural time series data. *J. Neurosci. Methods* **352**, 109080 (2021).
46. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Comput. Intell. Neurosci.* **2011**, 1–13 (2011).
47. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
48. Oosterhof, N. N., Connolly, A. C. & Haxby, J. V. CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. *Front. Neuroinform.* **10**, 1–27 (2016).
49. Taulu, S. & Simola, J. Spatiotemporal Signal Space Separation Method for Rejecting Nearby Interference in MEG Measurements. *Phys. Med. Biol.* **51**, 1759 (2006).
50. Guggenmos, M., Sterzer, P. & Cichy, R. M. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *Neuroimage* **173**, 434–447 (2018).
51. Gaser, C. & Dahnke, R. CAT-a computational anatomy toolbox for the analysis of structural MRI data. in *OHBM* 336–348 (2016).
52. Hämäläinen, M. S. & Ilmoniemi, R. J. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* **32**, 35–42 (1994).
53. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
54. Wurm, M. F. & Lingnau, A. Decoding actions at different levels of abstraction. *J. Neurosci.* **35**, 7727–7735 (2015).
55. Cross, E. S., Stadler, W., Parkinson, J., Schütz-Bosbach, S. & Prinz, W. The influence of visual training on predicting complex action sequences. *Hum. Brain Mapp.* **34**, 467–486 (2013).
56. Baldassano, C. *et al.* Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709-721.e5 (2017).
57. Silver, N. C. & Dunlap, W. P. Averaging correlation coefficients: Should Fisher's z transformation be used? *J. Appl. Psychol.* **72**, 146–148 (1987).
58. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**,

177–190 (2007).
59. Kiesel, A., Miller, J., Jolicœur, P. & Brisson, B. Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology* **45**, 250–274 (2008).
60. Smulders, F. T. Y. Simplifying jackknifing of ERPs and getting more out of it: Retrieving estimates of participants' latencies. *Psychophysiology* **47**, 387–392 (2010).

## Acknowledgments
We thank Danielle E. Parrott for kindly serving as the ballet dancer in our video stimuli. We thank Gianpiero Monittola and Davide Tabarelli for technical support in the MEG lab.

## Author contributions
Conceptualization: MFW
Methodology: IEJV, MFW
Investigation: IEJV, MFW
Visualization: IEJV, MFW
Supervision: MFW
Writing—original draft: IEJV
Writing—review & editing: IEJV, MFW

## Competing interests
The authors declare that they have no competing interests.

## Data and materials availability
All analysis scripts, the experiment script, and the stimuli used for the here presented results are freely available on GitHub at: https://github.com/Ingmar-de-Vries/DynamicPredictions. As the authors are still using this MEG dataset for a successive project, it is currently only available upon reasonable request, but will be made publicly available at a later stage.

# Supplementary material

| Stimulus | Ballet figures | | | |
|---|---|---|---|---|
| 1 | Bow left | Arabesque left | Pirouette | Passé back |
| 2 | Bow left | Passé left | Jump right | Pirouette |
| 3 | Bow front | Jump back | Passé front | Arabesque front |
| 4 | Pirouette | Bow front | Passé front | Jump right |
| 5 | Pirouette | Passé right | Bow front | Jump back |
| 6 | Passé right | Bow right | Arabesque right | Pirouette |
| 7 | Passé right | Pirouette | Jump left | Arabesque front |
| 8 | Passé front | Arabesque front | Bow left | Jump right |
| 9 | Jump left | Bow right | Arabesque front | Passé left |
| 10 | Jump right | Pirouette | Passé front | Arabesque right |
| 11 | Jump back | Arabesque front | Pirouette | Bow front |
| 12 | Arabesque left | Pirouette | Jump back | Passé front |
| 13 | Arabesque right | Passé right | Pirouette | Bow front |
| 14 | Arabesque right | Jump left | Bow right | pirouette |

**Table S1. Ballet figures per stimulus.** Each stimulus consisted of a unique sequence of 4 smoothly connected ballet figures. The 4 figures were selected from a total of 5 unique figures. Note that all figures were presented from multiple viewpoints, to allow for studying absolute body posture and motion separately from relative (i.e., orientation and position invariant) body posture and motion. The pirouette could start and end at different angles and was always performed counter-clockwise.

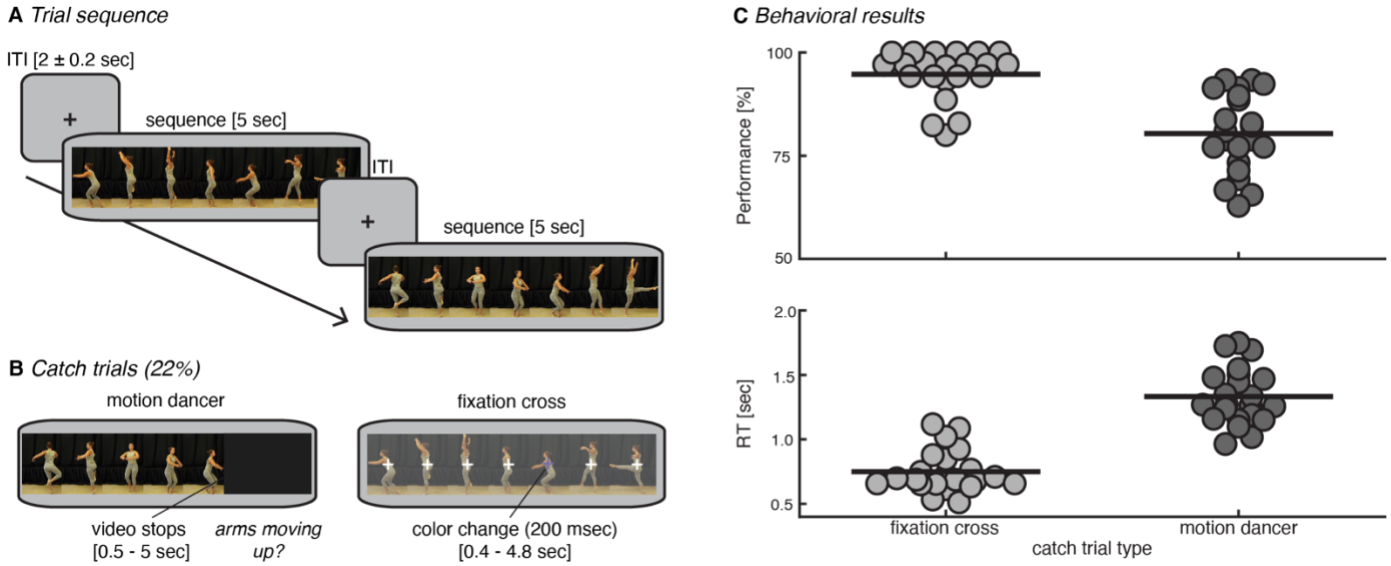


**Figure S1. Experimental design and behavioral results.** (a) Trial sequence. A run consisted of 90 sequences of 5 seconds alternated by blank screens with a fixation cross presented for 1.8 – 2.2 seconds (uniform distribution). (b) To ensure attention to the ballet dancer, 15 trials per run were randomly interrupted between 0.5–5 seconds after onset, and subjects were asked to (dis)confirm (2 response buttons) a statement about the particular ballet motion currently being performed (left panel). To stimulate fixation, on 5 trials per run the fixation cross changed color for 200 msec between 0.4 and 4.8 seconds after onset, and subjects had to press a button upon detection (right panel). (c) Behavioral results. Percentage correct and trial-averaged RT on catch trials in top and bottom panels, respectively. Dots represent single-subject data. Horizontal thick lines represent the group mean.

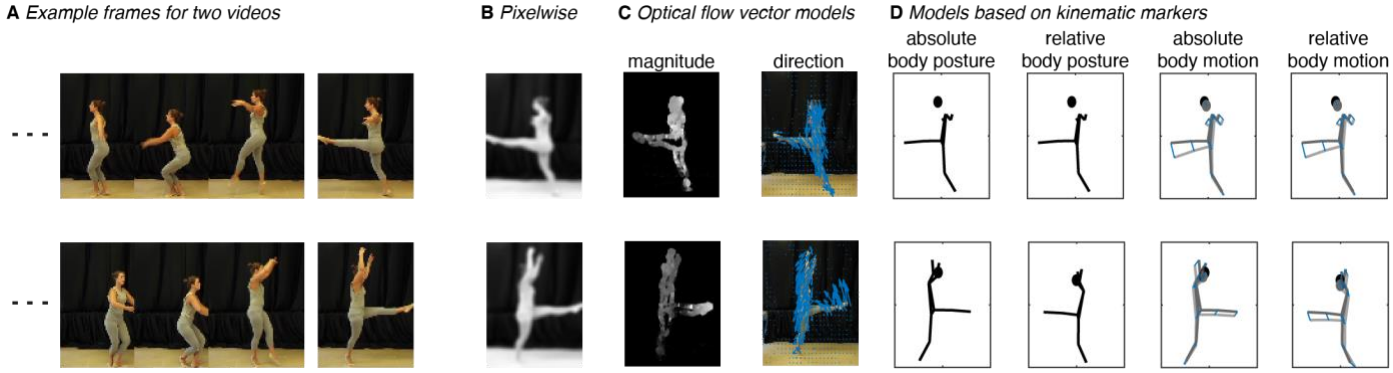


**Figure S2. Illustration of stimulus models.** (a) Example frames of two videos (rows) to be correlated for creating the model RDMs. From left to right; (b) the pixelwise smoothed grayscale of the video frame, (c) the magnitude of the optical flow vectors (brighter color is higher magnitude) and the direction of the optical flow vectors (as indicated by blue arrows, which are scaled 10 times here for illustrative purposes), and (d) absolute body posture (as defined by 3D kinematic marker positions), relative body posture (i.e., after aligning the kinematic markers of the second video with those of the first video without changing the internal structure, through translation and rotation along the vertical axis), absolute body motion as indicated by blue lines (i.e., difference in kinematic marker position between two subsequent frames), and relative body motion.

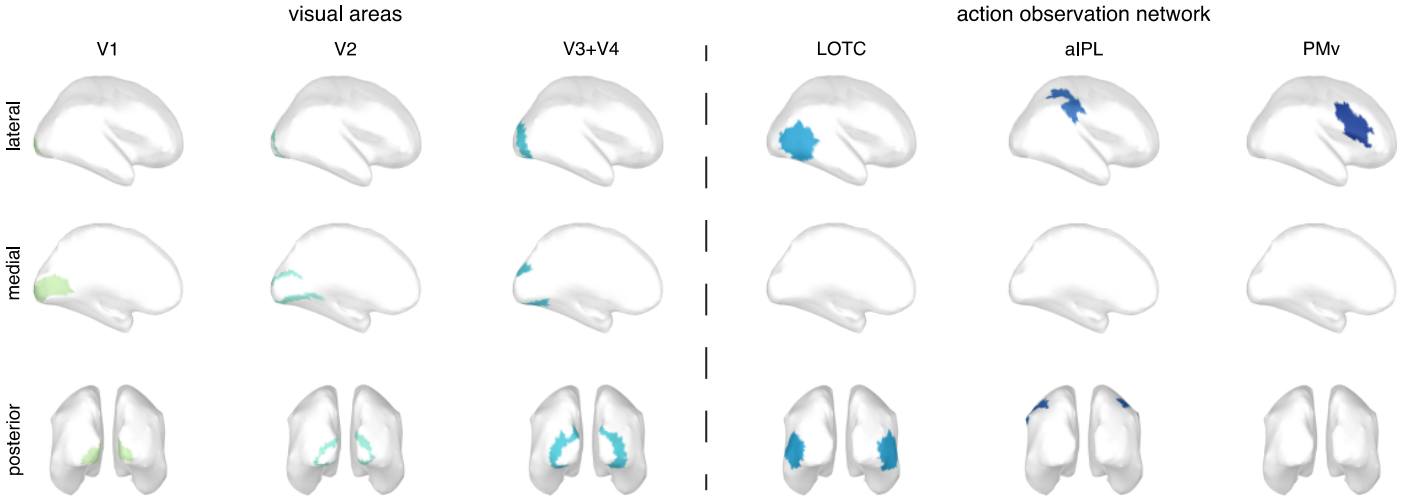


**Figure S3. ROI definitions.** Cortical regions of interest for main ROI analysis (Fig. 2), based on combinations of parcels as defined according to the Human Connectome Project (HCP) atlas [53]. Here ROIs were a priori defined as follows (ROI name = atlas parcels): V1 = V1 [487 vertices]; V2 = V2 [491 vertices]; V3+V4 = V3 and V4 [490 vertices]; LOTC = V4t, FST, MT, MST, LO1, LO2, LO3, PH, PHT, TPOJ2 and TPOJ3 [565 vertices]; aIPL = PF, PFt, AIP and IP2 [403 vertices]; PMv = IFJa, IFJp, 6r, 6v, PEF, IFSp, 44 and 45 [466 vertices]. Note that these vertex amounts are based on the ICBM152 template cortical surface, exact amounts differ slightly between individual subjects. MEG responses at all vertices from both hemispheres were combined into a single vector to compute the neural RDM at a single time point (i.e., pairwise dissimilarity in the MEG response to the 14 ballet sequences). For visualization, all vertices within a single ROI are given the same color, which matches the color for each ROI in Figure 2. LOTC = lateral occipitotemporal cortex, aIPL = anterior inferior parietal lobe, PMv = ventral premotor cortex.

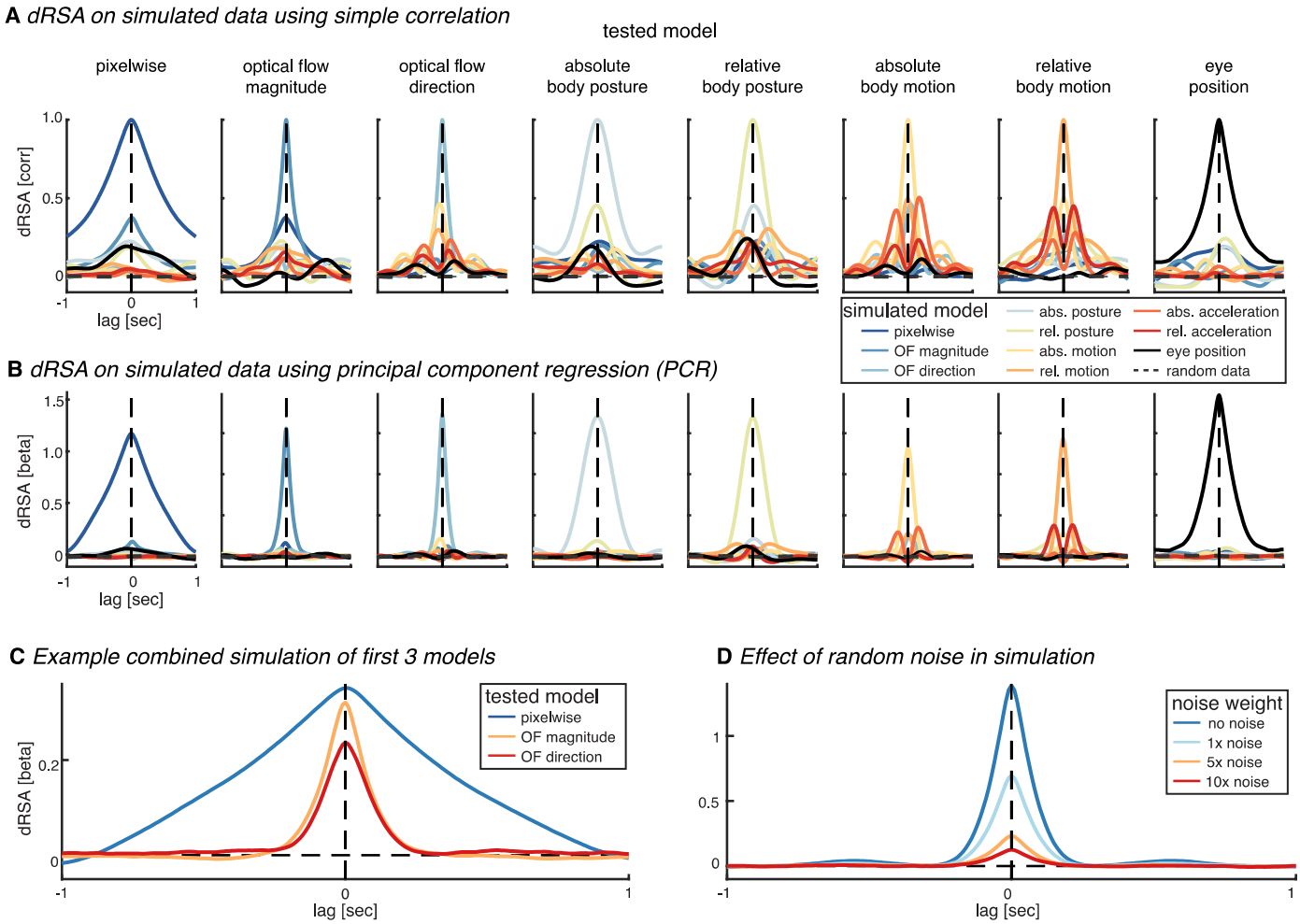


**Figure S4. Simulations.** (a) Correlation as similarity measure for dRSA. Colors indicate separate neural simulations of individual model RDMs. Columns indicate which model was tested. These results illustrate the effect of shared variance between various models and the effect of temporal autocorrelation within a given model. (b) Regression weight as similarity measure for dRSA using principal component regression (PCR; see "Materials and Methods" for details). In short, simulated neural RDMs are exactly the same as in (a), but at the final step in dRSA (i.e., comparing neural and model RDMs) all models are included in a single regression, and only the beta weight of the tested model (columns) is illustrated. Importantly, these results indicate that PCR is effective at extracting only the model of interest from the simulated neural RDM, while largely regressing out the other models. (c) The first 3 models (pixelwise, optical flow magnitude and direction) are combined to create a single simulated neural RDM. In each dRSA analysis, only a single model is

tested (colors), while the others are regressed out using PCR (as in b). Importantly, PCR is effective in extracting a single model while regressing out the other two. (d) Same PCR analysis as (b) and (c), with optical flow direction as simulated and tested model, but with different implant-to-random noise ratios to create the simulated neural RDM.

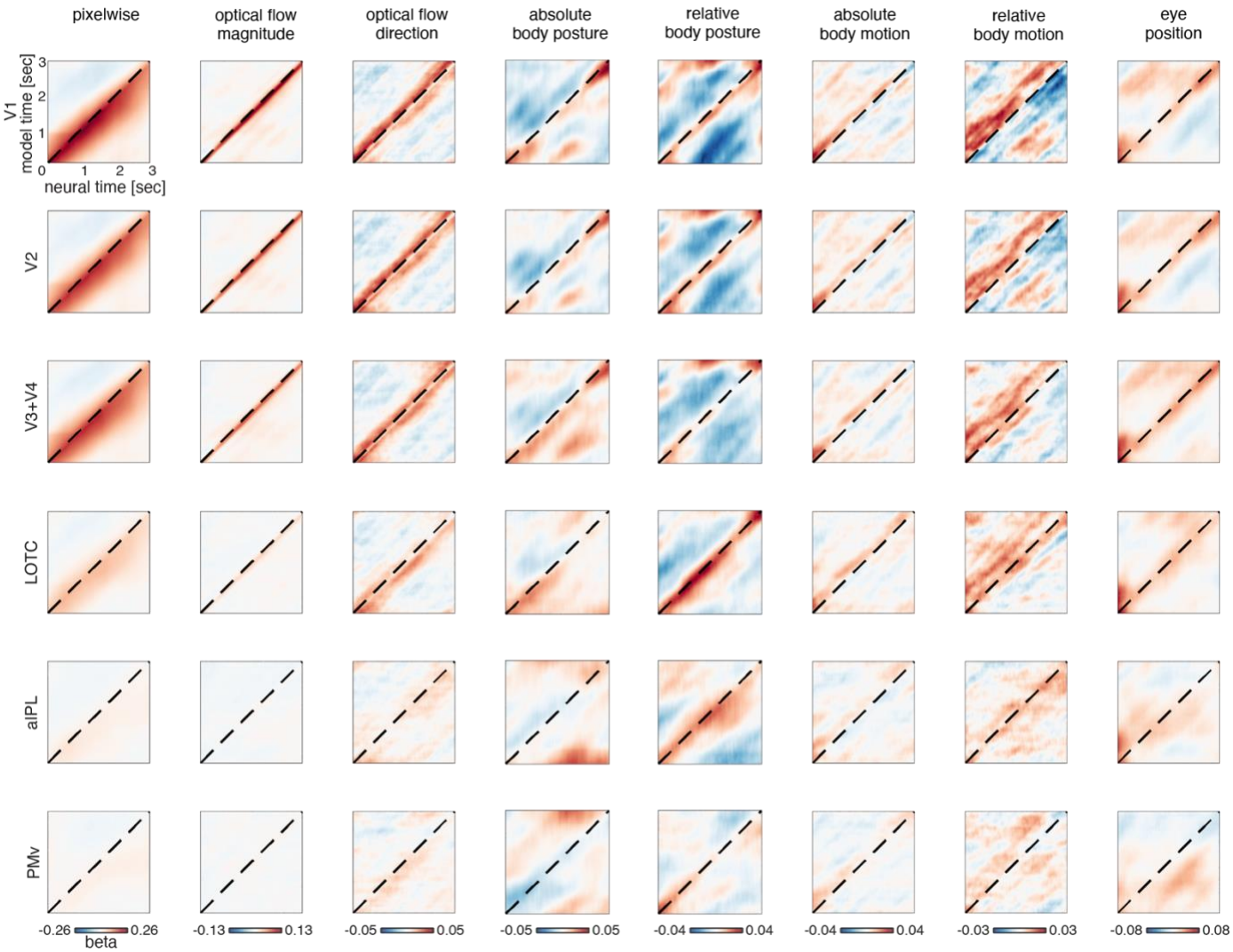


**Figure S5. Neural-time by model-time two-dimensional dRSA.** Dynamic RSA regression weights, illustrated as 2-dimensional model-time by neural-time dRSA matrix, where stronger weights in the lower right triangle indicate that the neural representation is lagging behind a model representation, while stronger weights in the upper left triangle indicate that the neural representation is preceding a model representation (i.e., a predictive neural representation). Different stimulus-feature models are plotted in columns, while ROIs are plotted in rows. The diagonal dashed line indicates zero lag between the neural and model RDM. Note that the scale illustrated below each column reflects the scale for that whole column.

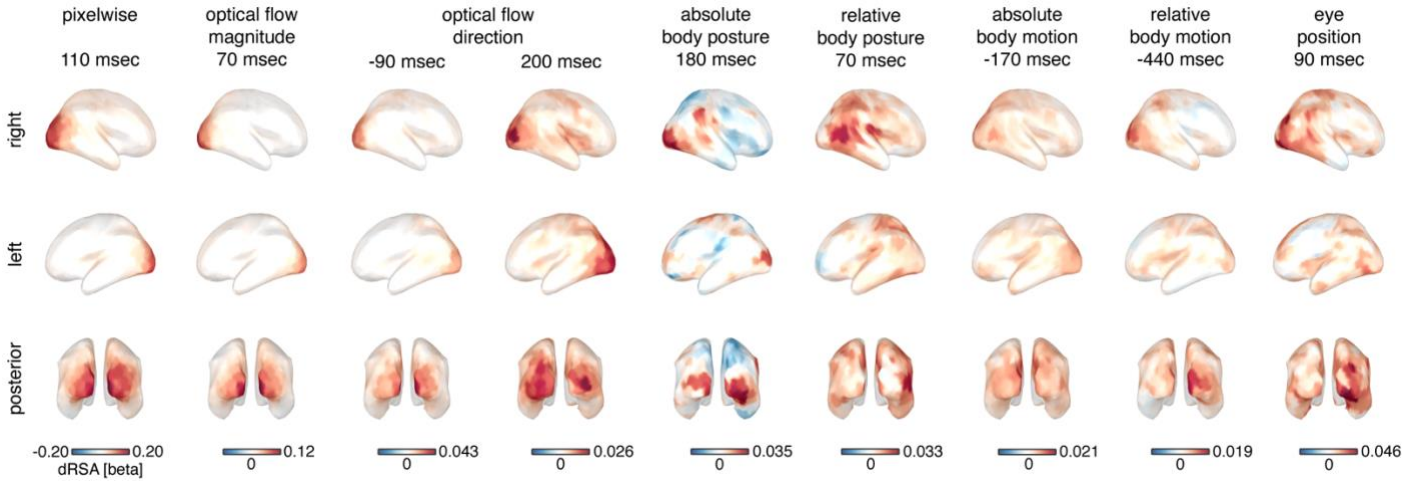


**Figure S6. Uncorrected searchlight dRSA on cortical surface.** Quasi-searchlight for which the dRSA analysis was run on all 400 parcels of the Schaefer atlas[24], which have roughly similar vertex amounts. Peak timings are selected per model based on the ROI analysis illustrated in Figure 2a. For FDR corrected source maps see Fig. 2b.