Predictive neural representations of naturalistic dynamic input

Ingmar E.J. de Vries*[1,2], Moritz F. Wurm[1]

[1] Centre for Mind/Brain Sciences (CIMeC), University of Trento, 38068 Rovereto, Italy.
[2] Donders Institute, Radboud University, 6525 EN Nijmegen, The Netherlands
* i.e.j.de.vries@gmail.com

Abstract:

Adaptive behavior requires our brain to predict unfolding external dynamics. While theories assume such dynamic prediction, empirical evidence is limited to static snapshots and indirect consequences of predictions. We present a dynamic extension to representational similarity analysis that uses temporally variable models to capture neural representations of unfolding events. We applied this approach to source-reconstructed MEG data of healthy human subjects, and demonstrate predictive neural representations of future posture and motion of observed actions. Predictive representations exhibit a hierarchical pattern, such that high-level abstract stimulus features are predicted earlier, while low-level visual features are predicted closer in time to the actual sensory input. This new approach offers a window into predictive processing of dynamic events, and can be applied to other sensory modalities and biosignals.

One-Sentence Summary:

Dynamic representational similarity analysis unveils how and when our brain represents and predicts dynamics of the world.

Main Text:

In dynamic environments, our brain is faced with a continuous stream of changing sensory input, which we use to update our expectations of unfolding events and react promptly. While theories assume such dynamic prediction[1–3], empirical research has focused on static snapshots and indirect consequences of predictions, often using simple stimuli of which predictability is manipulated[4–8]. However, our complex dynamic world necessitates us to continuously update predictions based on fast-changing external information. For example, to catch a ball, the brain needs to estimate its future trajectory (based on the opponent's intentions, the throwing movement and the ball's initial trajectory) in order to plan appropriate motor commands for catching the ball. How does the brain represent dynamic events at these different levels, and how do these representations temporally relate to (i.e., follow or precede) actual events in the real world? A powerful approach to investigate neural representations is representational similarity analysis (RSA[9]), which quantifies the similarity between neural and model representations. RSA has been mainly used to investigate neural representations of static stimuli (e.g. pictures of objects), using static models (e.g. for shape or category[10]). In order to investigate representational dynamics of temporally changing stimuli, we developed a dynamic extension to RSA that uses temporally variable models to capture neural representations of unfolding events (Fig. 1). In dynamic RSA, similarity between neural and model representations is computed at all neural-by-model time points. Hereby, dRSA allows investigating the match between a 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 state and the best-matching neural representation (i.e., the time needed for information to pass from retina to e.g., V1). In contrast, a negative lag should be observed for predictive neural representations, in which case neural representational content predicts the future model state. Here we applied dRSA to source-reconstructed MEG data of healthy human subjects observing dance videos that were modeled along various levels of the visual processing hierarchy (from low-level visual to view-dependent and -invariant body posture and motion). Attention to the unfolding events was confirmed by high performance on occasional test trials, in which subjects were tested on the dancer's motion, whereas eye fixation was ensured by subjects' successful detection of an occasional subtle color change of the fixation cross displayed in the center (Fig S1b and c).

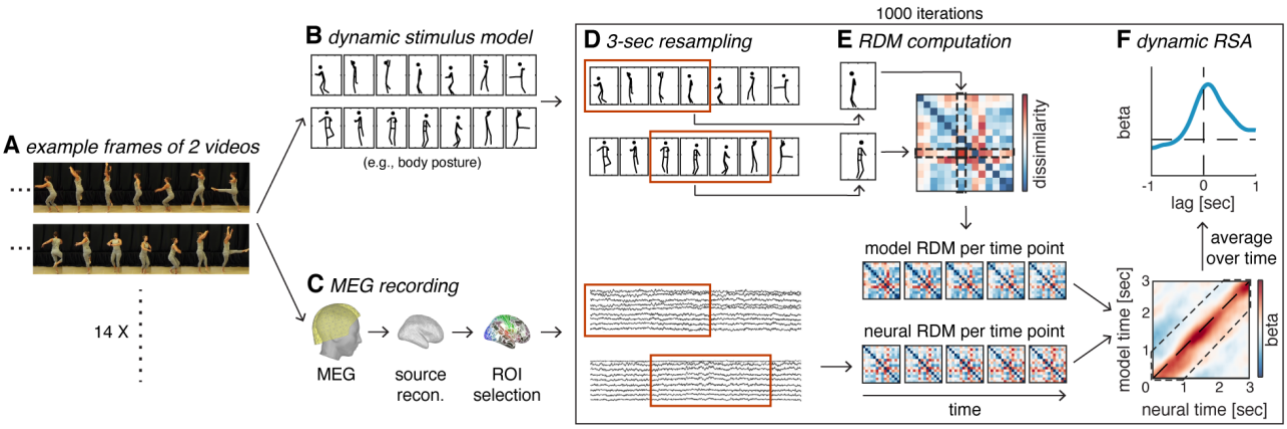

Figure 1. Dynamic RSA. (**A**) Subjects observed ~36 repetitions of 14 unique 5-sec dancing videos (Table S1 and Fig. S1a) during MEG. (**B**) Stimuli were characterized at different levels using dynamic stimulus models (e.g., body posture; Fig. S2). **(C)** Individual-subject source-reconstructed MEG signals within regions of interest (ROI; Fig. S3) were used as features for subsequent steps. (**D**) Temporal resampling was used to attenuate temporal heterogeneity of representational variance. Specifically, across 1000 iterations, a 3-sec window was randomly selected for each of the 14 stimuli. (**E**) Subsequently, on each iteration, neural and model representational

dissimilarity matrices (RDM) were created at each time point within the 3-sec window, based on pairwise dissimilarity in neural responses to the 14 stimuli and in stimulus feature models, respectively (here shown for 5 time points). (**F**) Last, similarity between neural and model RDMs was computed for each neural-by-model time point (lower panel), using regression weights to test a specific model RDM, while regressing out other covarying model RDMs. This approach was validated through simulations (Fig. S4). The 2-dimensional dRSA matrix was averaged along the diagonal to create a lag-plot (i.e., the lag between neural and model RDM; upper panel), in which peaks to the right or left of the vertical zero-lag midline reflect lagged or predictive neural representations, respectively. The result is subsequently averaged over resampling iterations for each subject, ROI and model, and statistically tested.

Dynamic RSA revealed distinct types of dynamic representations with unique characteristics (Fig. 2): Most importantly, body motion was represented in a broad temporal window preceding the actual input by ~640–430 msec for view-invariant body motion, and ~300–150 msec for view-dependent body motion (Fig. 2a and 3a), indicating that these neural representations predicted future body motion. A computer vision model capturing low-level motion as optical flow vector direction at each pixel was represented predictively at ~90 msec. Predictive representations were present in all regions of the action observation network (AON), albeit strongest in visual areas, which was confirmed by a whole-brain analysis (Fig. 2b). While several previous studies revealed the consequence of motion prediction in early visual cortex[8,11], our results provide a first characterization of how neural prediction of future motion in naturalistic stimuli continuously unfolds along the visual processing hierarchy. Specifically, peak latencies of predictive motion representations reflected the top-down order along the processing hierarchy, such that view-invariant body motion was predicted earliest in time, followed by view-dependent body motion and optical flow direction (Fig. 3a; main effect of model in a posthoc ANOVA with factors model and ROI; F = 32.33, p = 3.2e-9). It may be that predictions originate at a conceptual level of how the observed action will unfold, as best captured by view-invariant body motion. This subsequently translates into more concrete predictions, to which eventually the sensory input can be compared. Such hierarchical prediction is in line with empirical findings from different sensory modalities[1,12–15] and contexts such as social perception and action observation[16,17]. It supports hierarchical Bayesian brain theories such as predictive coding, which postulate that higher level predictions act on, and therefore must precede, lower level predictions[18–20].

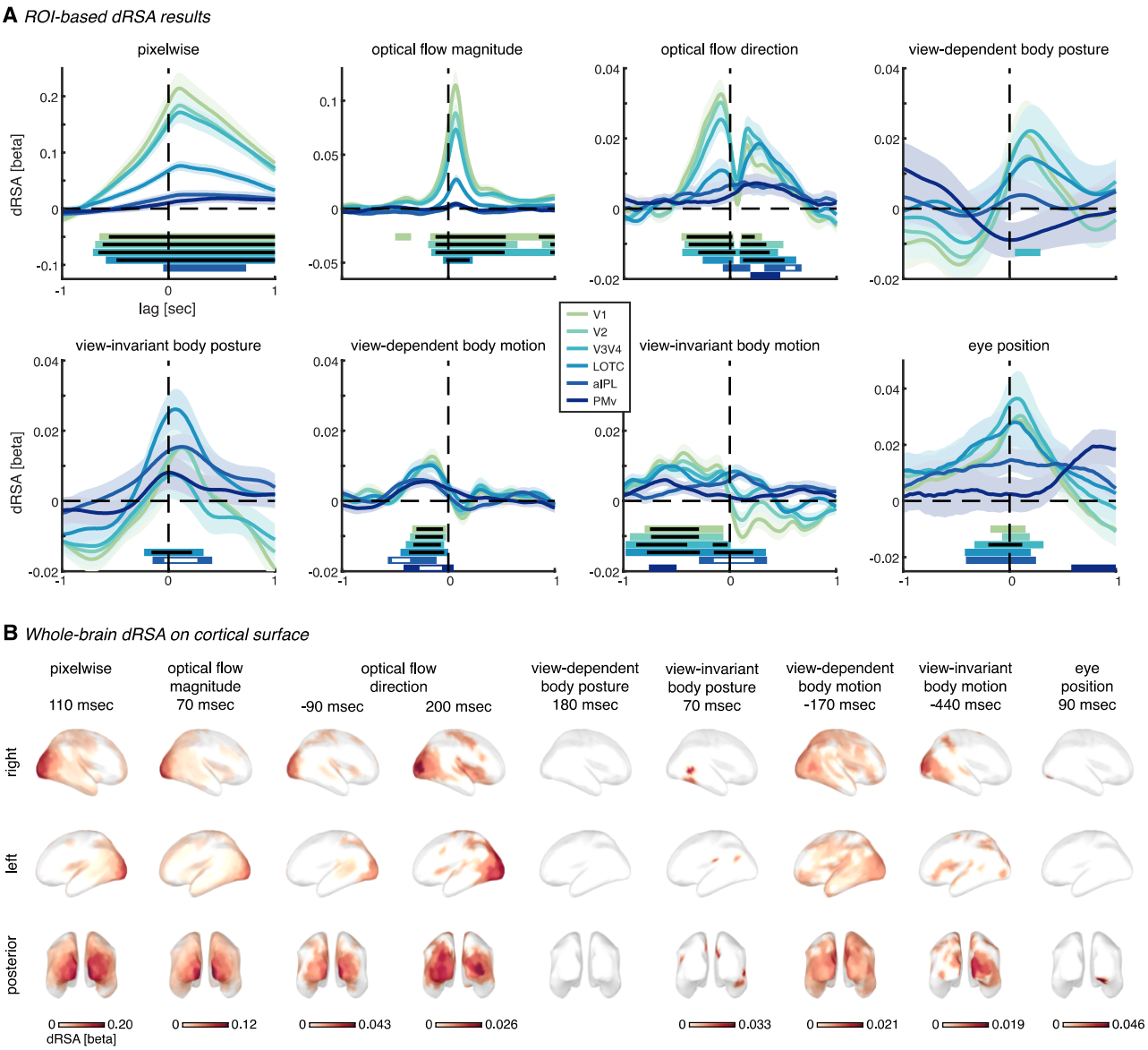

Figure 2. dRSA results. (**A**) ROI-based analysis. Dynamic RSA regression weights, illustrated as lag-plots (see Fig. S5 for 2D dRSA plots). Light-to-dark colors reflect posterior-to-anterior ROIs. Stimulus-feature models are plotted in separate subplots, with pixelwise capturing low-level visual information, optical flow magnitude and direction capturing low-level visual motion in each of the pixels, and body posture and motion models capturing 3D position and motion of 13 kinematic markers placed on the dancer (see Fig. S2 for details). Additionally, while subjects were instructed to keep fixation, an eye-position-based RDM was included to control for any remaining oculomotor activity (e.g., micro-saccades). 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 across time ($p < 0.05$ for cluster-size), with colors matching the respective ROI line plot. 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). (**B**) Whole-brain dRSA on cortical surface (FDR corrected at $p < 0.05$). Peak timings were selected per model based on the ROI analysis illustrated in (a).

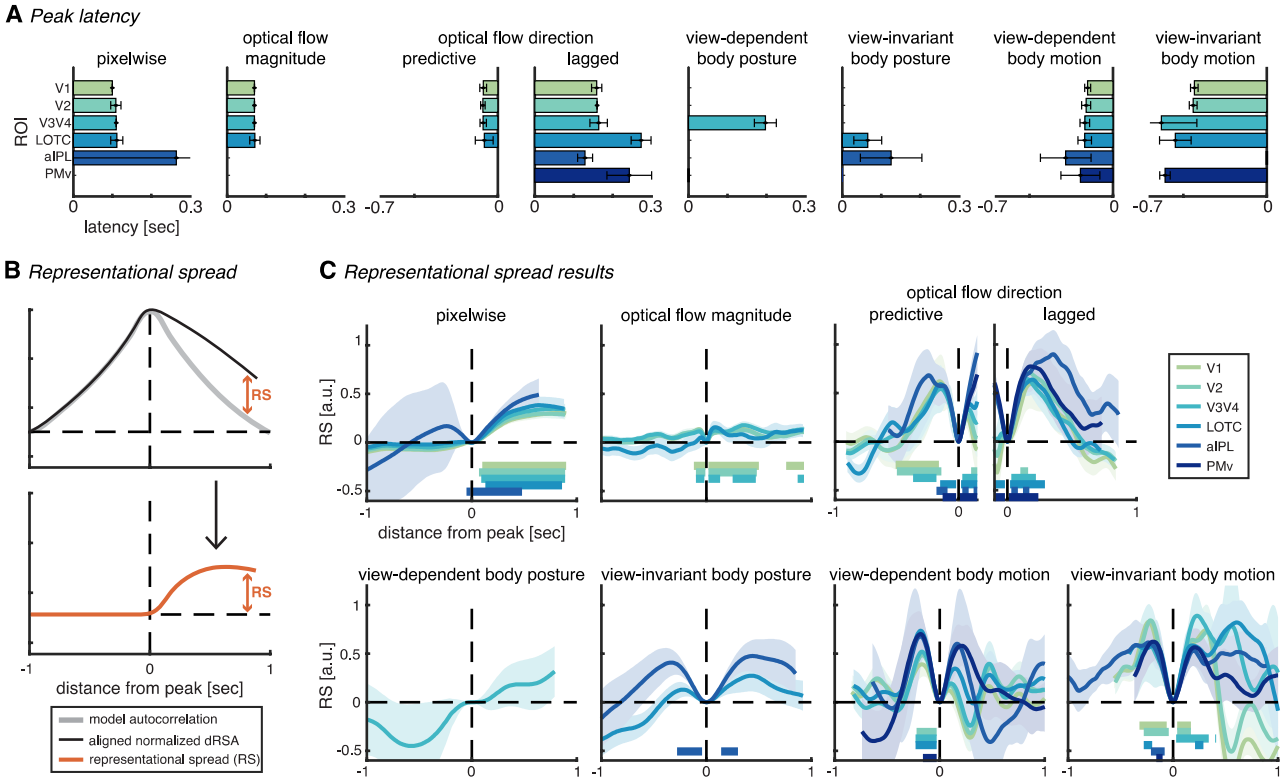

Figure 3. Peak latency and representational spread. Results are plotted for ROI-model combinations with a significant dRSA result (Fig. 2). (**A**) Peak Latencies of the dRSA from Fig. 2. Bars and error bars indicate jackknifed subject average and SEM, respectively. (**B**) Representational spread (RS), which indicates temporal information spread surrounding the dRSA peak, above and beyond what can be explained by model autocorrelation. To compute RS, we normalized both the observed dRSA curve and the model autocorrelation (Fig. S4b) by their peak value, aligned the curves, and subtracted model autocorrelation from observed curves. RS of 1 and 0 indicate maximally sustained and not sustained representation, respectively. The horizontal axis indicates temporal distance from the corresponding peak, with distance = 0 coinciding with the peak latency. Positive and negative temporal distances reflect representational spread on the right and left side of the dRSA peak, respectively. Since the dRSA peak latency determines the temporal window at which the RS can be computed, different RS lines end at different distances. (**C**) Representational spread results. Lines and shaded areas indicate jackknifed subject average and SEM, respectively. Same color conventions as Fig. 2, with a significance threshold of $p < 0.001$ for single samples and $p < 0.01$ for cluster-size. Note that as we observed both a predictive and lagged peak for optical flow direction (Fig. 2), we computed RS separately for each peak.

The representation of body posture was lagged, pointing towards bottom-up activation. However, we argue that this lagged representation might not necessarily reflect mere bottom-up processing, but instead is modulated by top-down prediction. That is, one would expect a higher-level representation such as view-invariant body posture to have a larger processing latency compared to low-level representations. By contrast, we observed a dRSA peak latency for view-invariant body posture in LOTC (~70 msec) that was earlier than the peak latency for the view-dependent body posture and pixelwise models (~180 and ~110 msec, resp.; Fig. 2a and 3a), suggesting that the observed neural representation of view-invariant body posture was modulated by top-down prediction. Thus, while a negative dRSA peak latency clearly evidences prediction, a positive latency precludes inferring that a representation is solely activated in a bottom-up manner. Prediction of body posture is in line with evidence for neural signatures of predictive processing of future object position[4,5,7]. The here identified prediction latencies likely depend on stimulus characteristics. Future work is needed to identify which factors (e.g., speed, predictability, prior knowledge) determine prediction latencies.

Interestingly, most lagged neural representations (i.e., pixelwise, optical flow magnitude and direction, and view-invariant body posture) were temporally sustained, indicating that representations were not immediately overwritten by new incoming visual input, but remained partially activated. We quantified the stability of representations by comparing the amount of information spread in the neural signal surrounding a dRSA peak with the amount that can be expected based on model autocorrelation alone (i.e., representational spread or RS; Fig. 3b and c). The fact that stimulus information is not immediately overwritten may provide a basis for neural predictions; that is, predicting the future requires integration of past information across an extended period for extrapolation[2]. As such, the sustained nature of representations might be necessary for creating and updating predictions based on constantly changing dynamic input. An alternative function of sustained representations might be the integration of features to form coherent higher-level representations. As different features have different processing times, an overlapping temporal window of representation is needed for their integration[2]. Future research should investigate how and which past information is integrated to form predictions, e.g., as shown for simple tone sequences[21] or by investigating the temporal characteristics of predictive representations after feature-specific perturbations. When additionally considering the peak latencies, representations of pixelwise vs. optical flow vector magnitude point towards a dissociation between representations drawing on parvocellular vs. magnocellular pathways, respectively: Pixelwise information was represented later (Fig. 3a; 100–110 msec) and more sustained (Fig. 3c), whereas optical flow vector magnitude was represented faster (Fig. 3a; 70–80 msec) and more transient (Fig. 3c). The difference in peak latencies was confirmed by a main effect of model in a posthoc ANOVA with factors model and ROI ($F = 33.69$, $p = 9.2e\text{-}6$). Such dissociation is in line with findings suggesting that the parvocellular pathway processes visual information at high spatial resolution with a slow and sustained response pattern, whereas the magnocellular pathway processes motion information with a faster and transient response pattern[22]. Interestingly, we also found representational spread preceding the peak latency (for body motion and optical flow direction), which might be explained by gradually, rather than abruptly, emerging predictive representations.

Taken together, dynamic RSA represents a methodological advancement for studying when our brain represents and predicts the dynamics of the world. It reveals both the peak latency and stability of neural representations of naturalistic dynamic input, and does so at several hierarchical levels of processing. In the current dataset it reveals hierarchical predictive representations of the future motion of an observed action, in support of Bayesian brain theories[18–20]. In principle, dRSA can be applied to any (naturalistic and controlled) dynamic stimulus (e.g., soundscape, music, language, motor planning/execution, social interaction) and any biosignal with high temporal resolution (e.g., M/EEG, EcoG, animal electrophysiology, EMG, eyetracking). Several recent studies investigated the dynamics of neural representations using related approaches. For instance, the temporal generalization approach investigates how neural activity patterns that discriminate between experimental conditions generalize across time[23]. Other approaches used RSA to study top-down processing by computing representational similarity between different brain regions at different time points to estimate the directionality of information flow[24], or used RSA to distinguish words prior to their presentation during naturalistic reading[25]. Importantly, all these approaches have in common that they use static models and/or stimuli. Dynamic RSA extends representational analysis methods to the study of dynamic events, thus paving the way for investigating the neural mechanisms underlying the representation and prediction of information across a wide range of processing levels in naturalistic dynamic scenarios.

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 stimulus set consisted of videos of 14 unique ballet dancing 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). We selected naturalistic body movements, rather than e.g., moving dots, because they are rich and evolutionary salient stimuli that are likely to evoke strong neural responses and that can be modeled across a wide range from low visual to view-dependent and -invariant levels of processing. Ballet dance was chosen to ensure comparably smooth body movements without abrupt action boundaries. All ballet figures were presented from multiple viewpoints, to allow for studying view-dependent body posture and motion separately from view-invariant 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 to each other (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 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 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 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 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 dance 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 the 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 dance 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[9]), 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[26] 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[27], 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.

View-dependent body posture (Fig S2d): We defined view-dependent 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.

View-invariant body posture (Fig. S2d): For the view-invariant 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 view-dependent body posture, but high similarity in terms of view-invariant 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.

View-dependent body motion (Fig. S2d): View-dependent body motion was operationalized as the first derivative of view-dependent 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.

View-invariant body motion (Fig. S2d): View-invariant body motion was computed the same way as view-dependent body motion, but after aligning one body posture to the other as described above for view-invariant 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.

View-dependent and view-invariant 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).

<u>MEG data collection and preprocessing</u>

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[29]. 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 offline using a combination of the Brainstorm[30], Fieldtrip[31] and CoSMoMVPA[32] toolboxes in MATLAB, as well as custom-written MATLAB scripts (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)[33] 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 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)[34], 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[35]. 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[36], 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)[37], 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)[38], and were expected to be important for processing and predicting these specific stimuli[17,39]. 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[32], 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., view-dependent and invariant 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 $91xN_{time}$ 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 $91xN_{time}$), 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 the signal-to-noise ratio. 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[40], 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 video stimuli (Fig. 1d), computed the RDMs (Fig. 1e) and dRSA (Fig. 1f) as described above on all 1000 iterations, and subsequently averaged dRSA matrices across iterations. While the 3-second windows were different for each video, they were aligned between stimulus model and corresponding neural data. With this random resampling procedure, the 14 videos are temporally realigned on each iteration, thus creating completely new RDM structures on each iteration. Note that this whole procedure is done on an individual subject level, and the averaged dRSA values are subsequently used for 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[41]. 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[42], 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 $91 \times N_{time}$ 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).

<u>Peak latency</u>

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[43]. 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[44]. 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, view-dependent body motion, and view-invariant 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 the temporal autocorrelation of models (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, 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[45], 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 and Notes

1. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).
2. Hogendoorn, H. Perception in real-time: predicting the present, reconstructing the past. *Trends Cogn. Sci.* **26**, 128–141 (2022).
3. Bubic, A., Yves von Cramon, D. & Schubotz, R. I. Prediction, cognition and the brain. *Front. Hum. Neurosci.* **4**, 1–15 (2010).
4. 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).
5. 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).
6. 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).
7. 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).
8. Aitken, F., Turner, G. & Kok, P. Prior Expectations of Motion Direction Modulate Early Sensory Processing. *J. Neurosci.* **40**, 6389–6397 (2020).
9. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 1–28 (2008).
10. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455–462 (2014).
11. 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).
12. Heilbron, M., Armeni, K., Schoffelen, J. M., Hagoort, P. & De Lange, F. P. A hierarchy of linguistic predictions during natural language comprehension. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–12 (2022).
13. Gayet, S. & Peelen, M. V. Preparatory attention incorporates contextual expectations. *Curr. Biol.* **32**, 687-692.e6 (2022).
14. 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).
15. 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).

16. 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).

17. Kilner, J. M., Friston, K. J. & Frith, C. D. Predictive coding: An account of the mirror neuron system. *Cogn. Process.* **8**, 159–166 (2007).

18. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).

19. 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).

20. Bastos, A. M. *et al.* Canonical Microcircuits for Predictive Coding. *Neuron* **76**, 695–711 (2012).

21. Baumgarten, T. J. *et al.* Neural integration underlying naturalistic prediction flexibly adapts to varying sensory input rate. *Nat. Commun.* **12**, 1–14 (2021).

22. Merigan, W. H. & Maunsell, J. H. R. How Parallel are the Primate Visual Pathways? *Annu. Rev. Neurosci.* **16**, 369–402 (1993).

23. King, J. R. & Dehaene, S. Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn. Sci.* **18**, 203–210 (2014).

24. 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).

25. 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).

26. 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).

27. 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

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. 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).

30. 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).

31. 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).

32. 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).

33. Taulu, S. & Simola, J. Spatiotemporal Signal Space Separation Method for Rejecting Nearby Interference in MEG Measurements. *Phys. Med. Biol.* **51**, 1759 (2006).

34. Guggenmos, M., Sterzer, P. & Cichy, R. M. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *Neuroimage* **173**, 434–447 (2018).

35. Gaser, C. & Dahnke, R. CAT-a computational anatomy toolbox for the analysis of structural MRI data. in *OHBM* 336–348 (2016).

36. Hämäläinen, M. S. & Ilmoniemi, R. J. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* **32**, 35–42 (1994).

37. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).

38. Wurm, M. F. & Lingnau, A. Decoding actions at different levels of abstraction. *J. Neurosci.* **35**, 7727–7735 (2015).

39. 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).

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

41. Silver, N. C. & Dunlap, W. P. Averaging correlation coefficients: Should Fisher's z transformation be used? *J. Appl. Psychol.* **72**, 146–148 (1987).

42. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

43. 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).

44. Smulders, F. T. Y. Simplifying jackknifing of ERPs and getting more out of it: Retrieving estimates of participants' latencies. *Psychophysiology* **47**, 387–392 (2010).

45. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).

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. We thank Christoph Huber-Huber for useful feedback on a previous version of the manuscript.

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:

Authors declare that they have no competing interests.

Data and materials availability:

All analysis scripts, the experiment script, and stimuli used for the here presented results are freely available 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. However, as soon as the authors are finished with the dataset, it will be made publicly available.

Supplementary Materials

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 view-dependent body posture and motion separately from view-invariant body posture and motion. The pirouette could start and end at different angles but was always performed counter-clockwise.

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

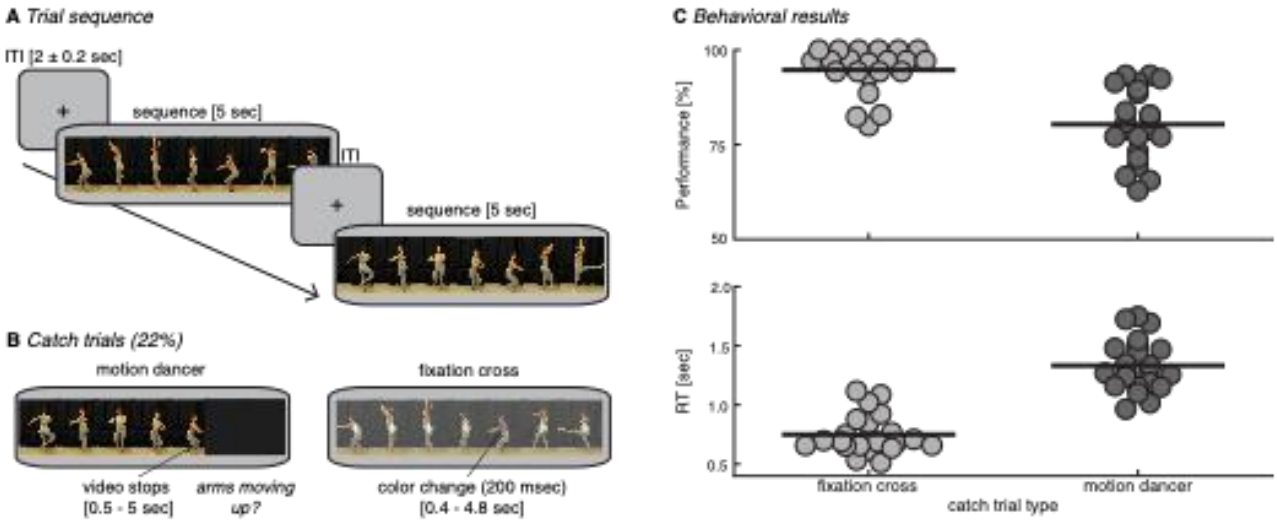

Figure S1. Experimental design and behavioral results. (**A**) Trial sequence. A run consisted of 90 sequences of 5 seconds alternated by blank screens with a fixation cross presented for 1.8 – 2.2 seconds (uniform distribution). (**B**) To ensure attention to the dancer, 15 trials per run were randomly interrupted between 0.5–5 seconds after onset, and subjects were asked to (dis)confirm (2 response buttons) a statement about the particular 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.

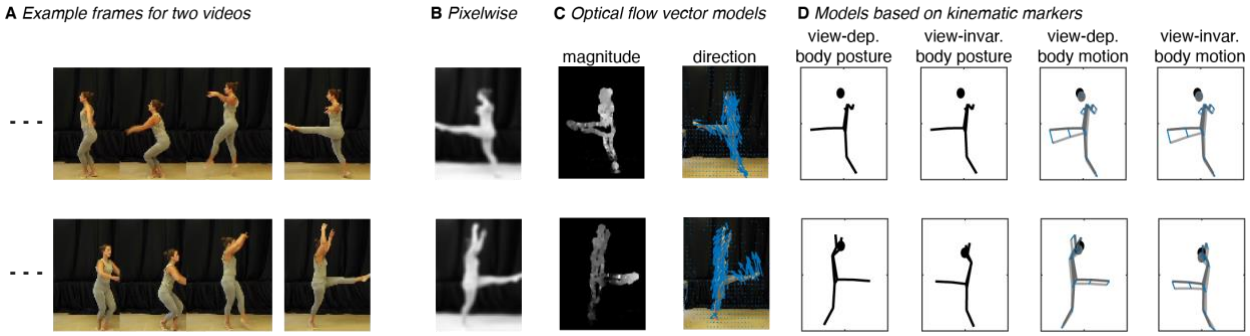

Figure S2. Illustration of stimulus models. (**A**) Example frames of two videos (rows) to be correlated for creating the model RDMs. From left to right; (**B**) the pixelwise smoothed grayscale of the video frame, (**C**) the magnitude of the optical flow vectors (brighter color is higher magnitude) and the direction of the optical flow vectors (as indicated by blue arrows, which are scaled 10 times here for illustrative purposes), and (**D**) view-dependent body posture (as defined by 3D kinematic marker positions), view-invariant 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), view-dependent body motion as indicated by blue lines (i.e., difference in kinematic marker position between two subsequent frames), and view-invariant body motion.

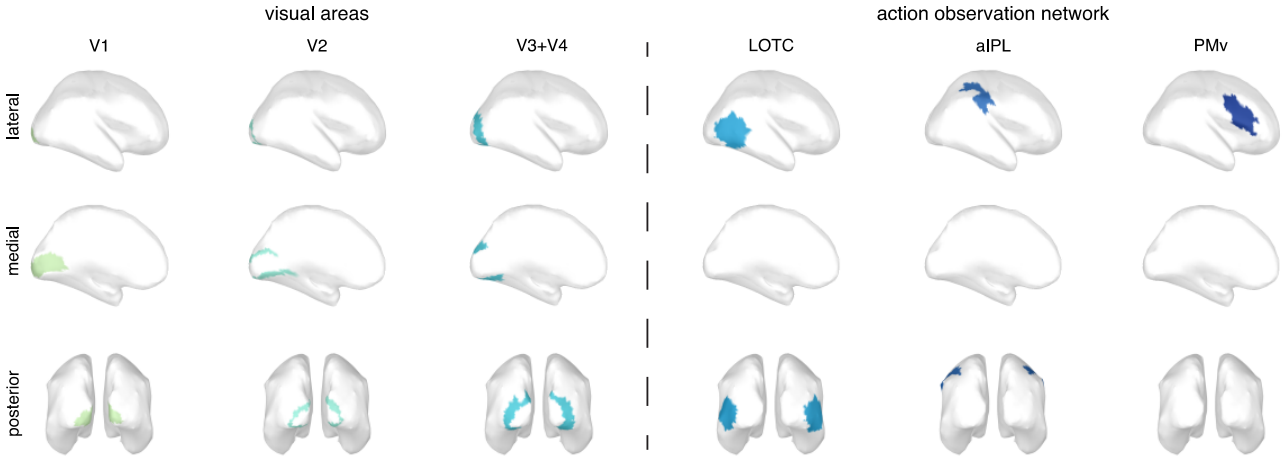

Figure S3. ROI definitions. Cortical regions of interest for main ROI analysis (Fig. 2), based on combinations of parcels as defined according to the Human Connectome Project (HCP) atlas (*37*). Here ROIs were a priori defined as follows (ROI name = atlas parcels): V1 = V1 [487 vertices]; V2 = V2 [491 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 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.

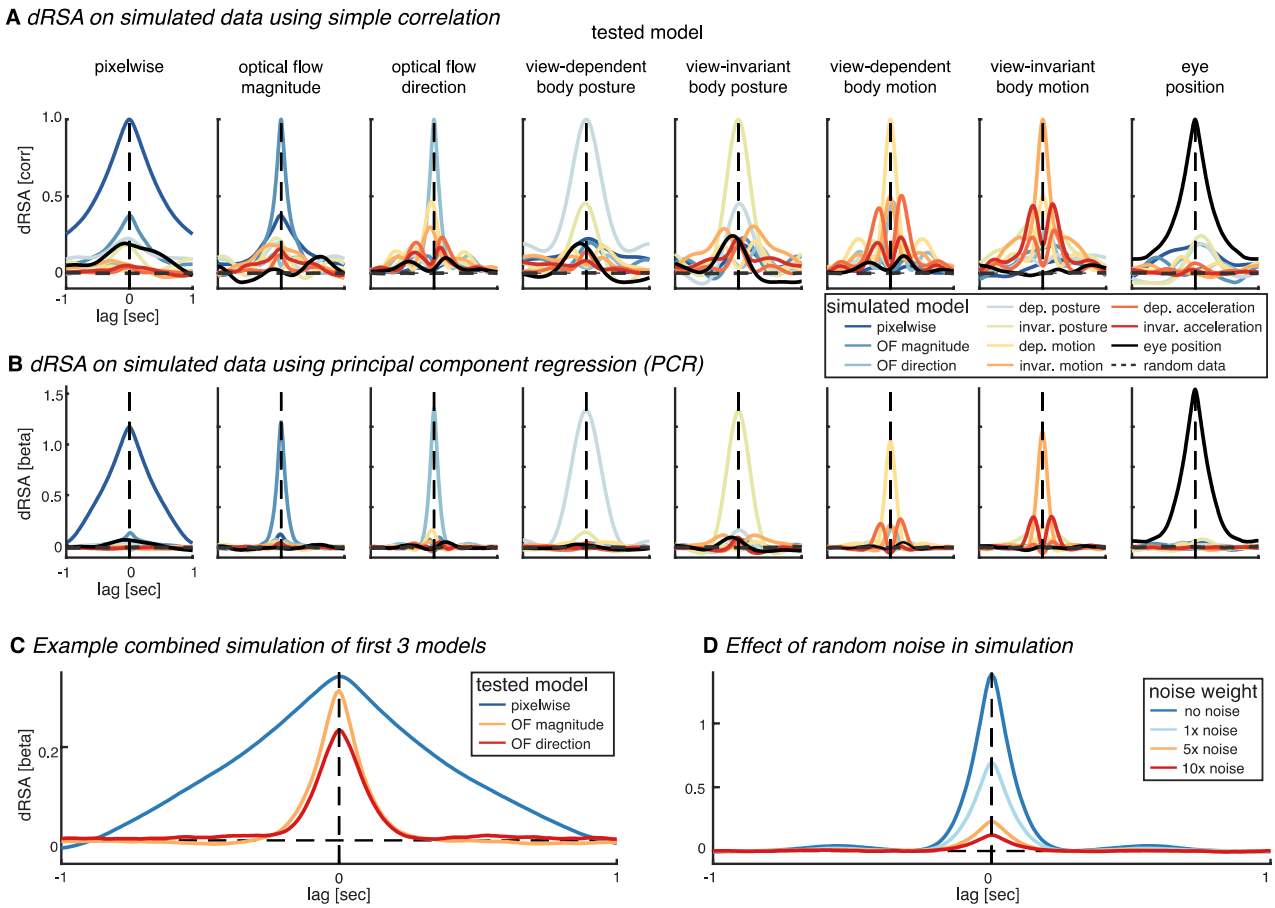

Figure S4. Simulations. (**A**) Correlation as similarity measure for dRSA. Colors indicate separate neural simulations of individual model RDMs. Columns indicate which model was tested. These results illustrate the effect of shared variance between various models and the effect of temporal autocorrelation 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.

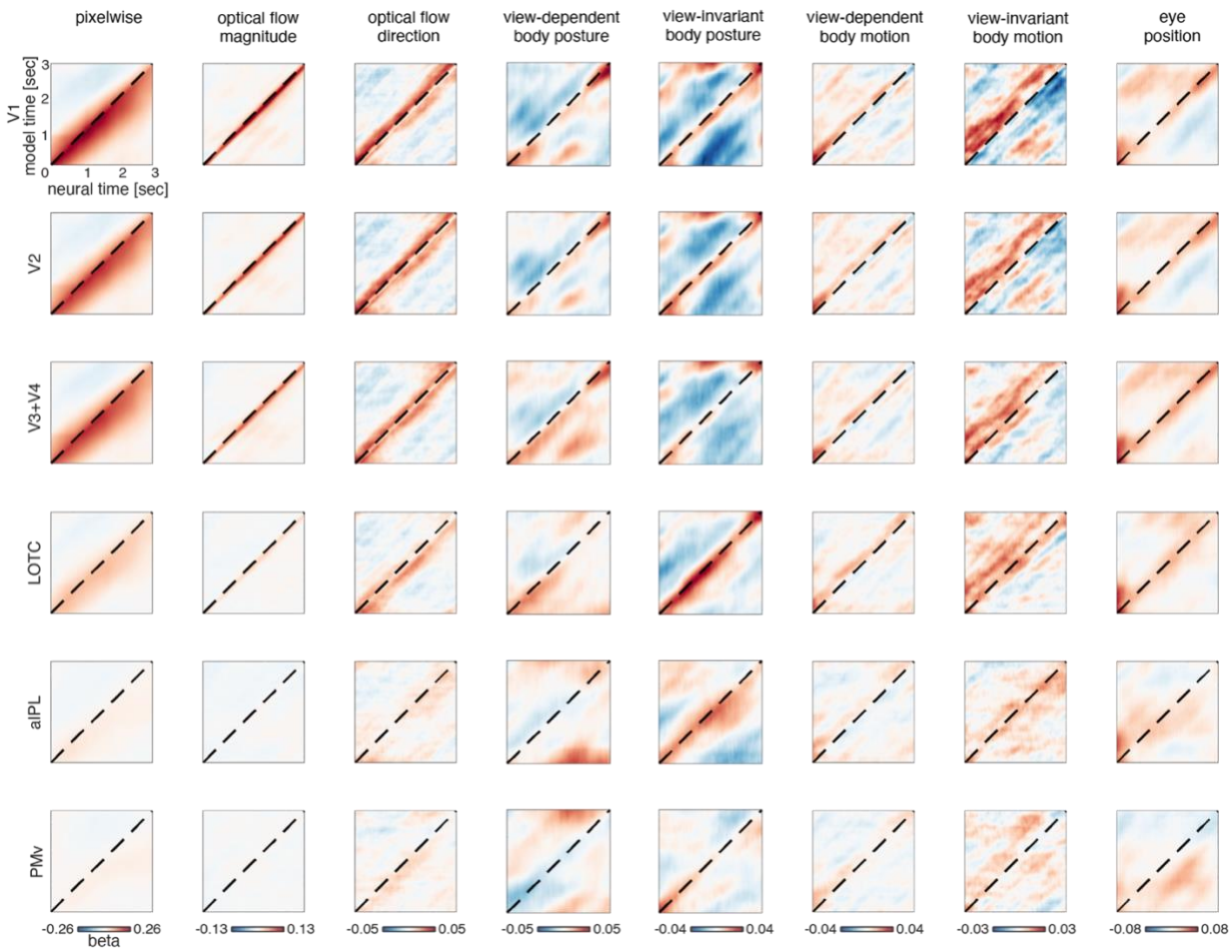

Figure S5. Neural-time by model-time two-dimensional dRSA. Dynamic RSA regression weights, illustrated as 2-dimensional model-time by neural-time dRSA matrix, where stronger weights in the lower right triangle indicate that the neural representation is lagging behind a model representation, while stronger 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.