Predictive neural representations of naturalistic dynamic input

Short title: Dynamic representational similarity analysis

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 such as social interaction 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 magnetoencephalography (MEG) data of healthy human subjects and demonstrate both lagged and predictive neural representations of observed actions. Predictive representations exhibit a hierarchical pattern, such that high-level abstract stimulus features are predicted earlier in time, while low-level visual features are predicted closer in time to the actual sensory input. By quantifying the temporal forecast window of the brain, this approach allows investigating predictive processing of our dynamic world. It can be applied to other naturalistic stimuli (e.g., film, soundscapes, music, motor planning/execution, social interaction) and any biosignal with high temporal resolution.

Introduction

In dynamic environments such as social interaction (e.g., traffic or sports), our brain is faced with a continuous stream of changing sensory input. Effective and prompt interaction with such environment requires our brain to predict unfolding external dynamics. While theories assume such prediction[1–3], empirical research has focused on static snapshots and indirect consequences of predictions, often using simple static stimuli of which predictability is manipulated[4–8]. However, in naturalistic dynamic environments, our brain needs to continuously update predictions of unfolding events for effective behavior. 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) 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 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]. 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. However, note that 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. That is, even post-stimulus representations are likely to be modulated by top-down expectations[11], for instance by sharpening the representation[12], or by shortening the processing latency[13,14]. Further note that as moment-by-moment stimulus predictability is determined by stimulus-specific feature trajectories and event boundaries, the resulting dRSA peak reflects the average latency at which a feature is represented most strongly for these stimuli. It does not mean this feature is exclusively represented at that exact latency.

Here we applied dRSA to source-reconstructed magnetoencephalography (MEG) data of healthy human subjects observing dance videos that were modeled at various levels of abstraction across the visual processing hierarchy, from low-level visual (i.e., pixelwise grayscale) to higher-level, perceptually more invariant (i.e., 3-dimensional view-dependent and -invariant body posture and motion), as to capture a comprehensive characterization of the dance sequences. Attention to the unfolding actions 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 (see "*Experimental design*"). Using dynamic RSA, we reveal the peak latency and spread of both lagged and predictive neural representations of naturalistic dynamic input, at several hierarchical levels of processing. By quantifying the temporal forecast window, dynamic RSA represents a methodological advancement for studying when our brain represents and predicts the dynamics of the world.

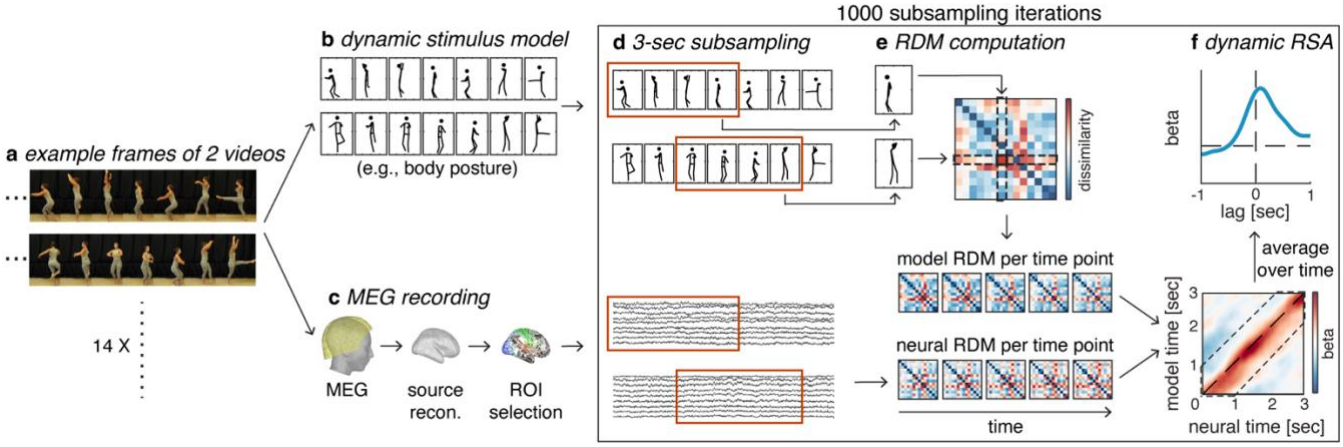


Figure 1. Dynamic RSA. a Subjects observed ~36 repetitions of 14 unique 5-sec dancing videos (Table 1 and Fig. 4a) during MEG. **b** Stimuli were characterized at different levels using dynamic stimulus models (e.g., body posture; see Fig. 5 for all models). **c** Individual-subject source-reconstructed MEG signals within regions of interest (ROI; see Fig. 6 for ROI definitions) were used as features for subsequent steps. **d** Temporal subsampling and realignment was used to attenuate idiosyncratic temporal heterogeneity in dRSA results caused by arbitrary pairwise alignment specific to these 14 stimuli (see subsection "*Temporal subsampling"*). In short, across 1000 iterations, a 3-sec segment was randomly extracted from each of the 14 5-sec stimuli (orange box), after which these new 14 subsampled 3-sec segments were realigned. Crucially, while a different random 3-sec window was selected for each of the 14 stimuli, for a given stimulus the temporal alignment between neural signal and models remained intact. **e** Subsequently, on each iteration, neural and model representational dissimilarity matrices (RDM) were created at each time point in the realigned 3-sec segments, based on pairwise dissimilarity in neural responses to the 14 stimuli and pairwise dissimilarity 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 (see subsection *"Simulations"* and Fig. 8). Last, 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. These dRSA lag-plots are computed within each subsampling iteration, after which they are averaged over iterations separately for each subject, ROI, and model, and statistically tested.

Results

Hierarchical motion prediction

Dynamic RSA revealed distinct types of dynamic representations with unique characteristics (Fig. 2a): Most importantly, body motion was represented in a broad temporal window preceding the actual input by ~500 msec for view-invariant body motion, and ~200 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 ~120 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[8,15], our results provide a first characterization of how neural prediction of future motion in naturalistic stimuli continuously unfolds at several timescales along the visual processing hierarchy. Specifically, peak latencies of predictive motion representations reflected the 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 vector direction (Fig. 3a; main effect of model in a post-hoc ANOVA with factors model and ROI; $F = 19.9$, $p = 8.3e-7$). Note that this does not evidence a strict serial cascade of prediction, such that low-level prediction is only enabled by earlier higher-level prediction. Instead, some low-level prediction likely remains even without any high-level prediction, as is apparent from the exclusively low-level prediction of simple motion stimuli that do not have a complex naturalistic hierarchical structure[16,17]. The three levels of prediction might therefore partly reflect independent streams. However, the temporal order of predictive representations observed here is in line with hierarchical Bayesian brain theories such as predictive coding that postulate that higher-level predictions act on, and therefore must precede, lower-level predictions[18–20]. Furthermore, such hierarchical prediction is also suggested by empirical findings from different sensory modalities[1,21–24], and complex contexts such as social perception and action observation[25,26]. In complex naturalistic stimuli as utilized here one would similarly expect a causal relationship amongst the partly independent prediction streams, such that high-level prediction affects low-level prediction. It may be that predictions originating at a higher, perceptually invariant level, as best captured by view-invariant body motion, subsequently modulate more concrete predictions, to which eventually the sensory input can be compared. Such causal relationship between different levels of prediction in naturalistic dynamic stimuli remains to be demonstrated in future research, for instance by selectively perturbing high-level stimulus predictability. Two expectations would be a shortening of the temporal forecast window for lower-level features (i.e., a less predictive dRSA peak), and a decrease in the strength of prediction.

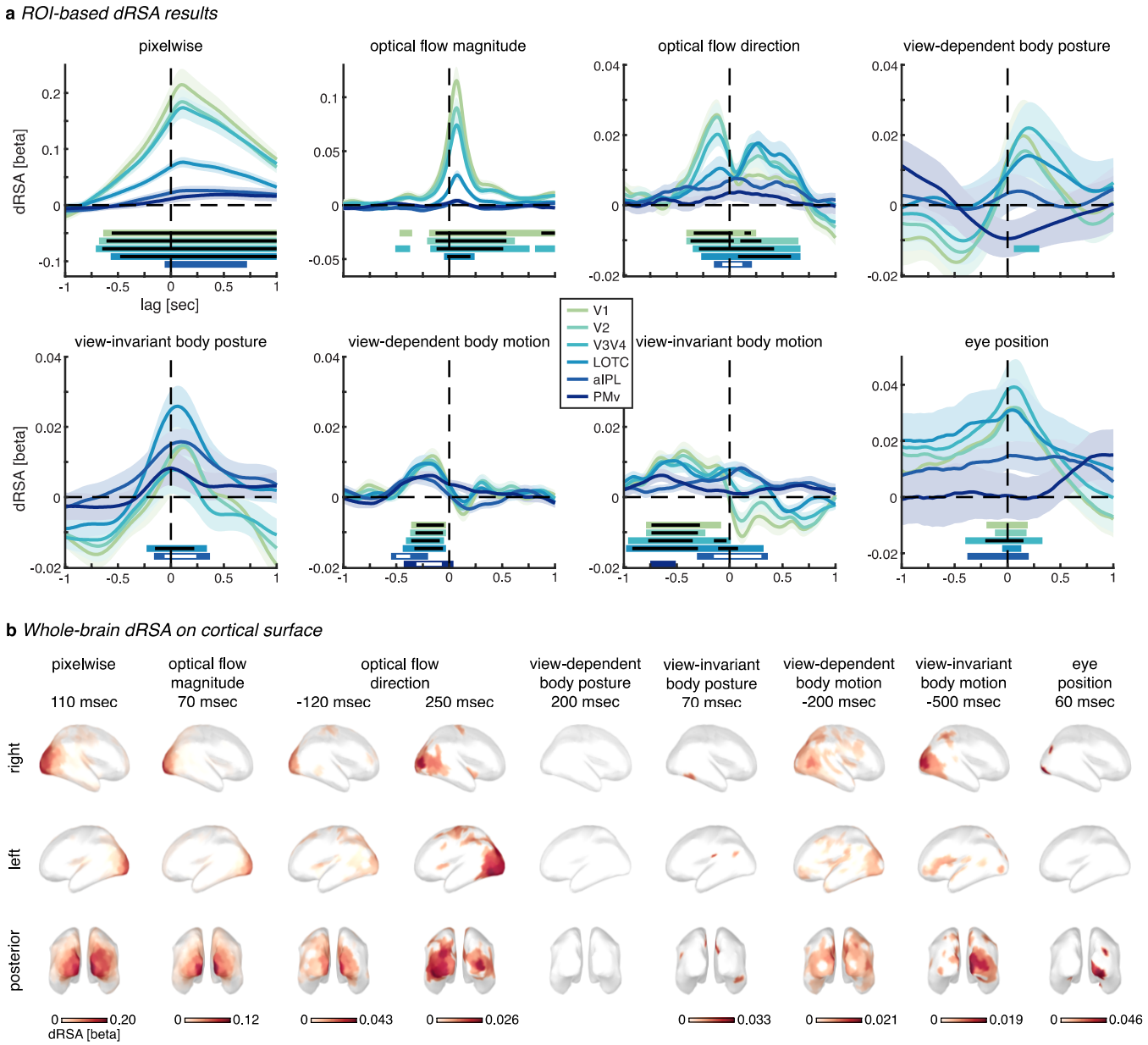


Figure 2. dRSA results. a ROI-based analysis. Dynamic RSA regression weights, illustrated as lag-plots (see Supplementary Fig. 1 for 2D dRSA plots). Light-to-dark colors reflect posterior-to-anterior ROIs (see subsection "*Source reconstruction and ROI selection*"). Stimulus-feature models are plotted 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 subsection "*Stimulus models*"). 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). See Fig. 6 for a comparison with the ROIs used in **a**.

Interestingly, for low-level motion (i.e., optical flow direction) we did not only observe a predictive, but also a lagged post-stimulus representation. According to predictive processing theories, post-stimulus representations should reflect the difference between prediction and sensory input (i.e., unpredicted input or "prediction errors")[27–31]. As such, it may be surprising that such post-stimulus representation is absent for high-level (view-dependent and -invariant) body motion. However, predictive processing theories hypothesize that accurate predictions effectively silence (or "explain away") all sensory input[27–31]. Albeit speculative, it may thus be that due to the highly predictable nature of smooth biological motion as in the ballet stimuli used here, motion information is effectively silenced after

the initial comparison at the lowest level, thus explaining the absence of lagged representation of high-level motion models. Such efficient silencing of redundant stimulus information might be particularly important in dynamic stimuli, as new potentially relevant input is continuously incoming. However, if accurate prediction has indeed silenced all sensory input in terms of high-level motion models, it is unclear why this does not hold for the posture models that clearly have lagged post-stimulus representations (see next section). Additionally, even for high-level motion information perfect prediction seems unlikely, and it is worth considering why dRSA might fail to pick up any remaining unpredicted information for high-level motion models. Note that dRSA reflects an average representation across stimulus time, stimuli, and subjects, rather than a single exclusive representation (latency). The little remaining unpredicted information may therefore simply be too variable in latency to be picked up in the average representation. Also, the signal-to-noise ratio (SNR) might generally be lower for bottom-up representations, as the visual cortex receives a much denser network of feedback relative to feedforward projections[31,32]. In any case, these results do not provide conclusive evidence for predictive processing/coding theory (i.e., that accurate predictions silence all sensory input[31]) but leave the door open for related theories that do hypothesize lagged post-stimulus representations even after accurate predictions such as adaptive resonance[33], or Bayesian inference without predictive coding[34]. Future research using dRSA should selectively perturb stimulus predictability at different hierarchical levels as described above to arbitrate between these theories.

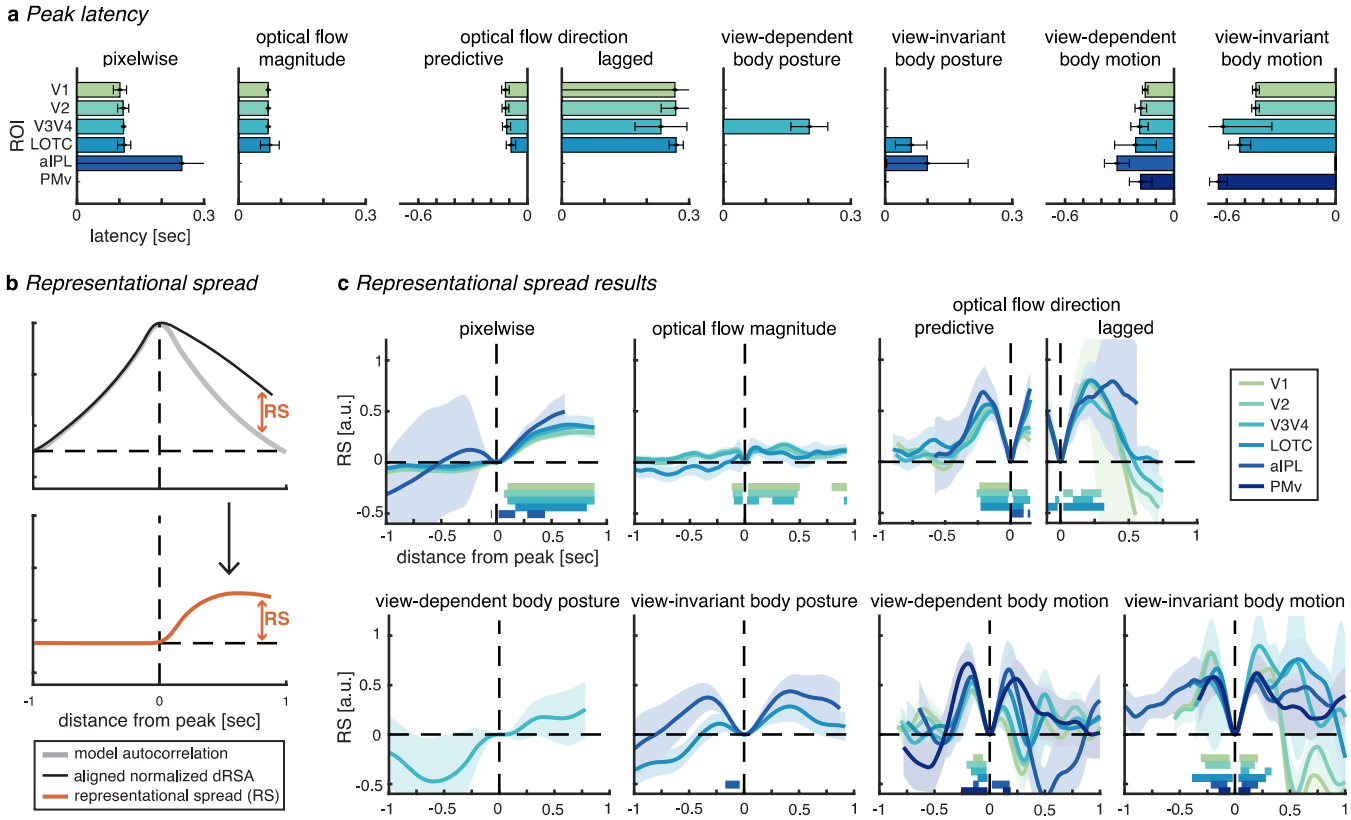


Figure 3. Peak latency and representational spread. Results are plotted for ROI-model combinations with a significant dRSA result (Fig. 2). **a** Peak Latencies of the dRSA from Fig. 2. Bars and error bars indicate subject average and SEM over retrieved jackknifed values, respectively (see subsection "*Peak Latency*"). Note that the variance indicated by the error bars reflects variance over retrieved jackknifed values, rather than individual-subject variance[35,36]. **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 (i.e., what remains after PCR; Fig. 8b) 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 averages and SEM, respectively. Same color conventions as Fig. 2a, 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. 2a), we computed RS separately for each peak.

Prediction modulates body posture representation

We observed lagged representations of view-dependent (~200 msec; V3/4) and view-invariant (~70 msec; LOTC and aIPL) body posture, which might suggest bottom-up activation. However, remember that a positive latency precludes inferring that a representation solely reflects bottom-up activation, as top-down expectations might very well modulate its veridity[12] and processing latency[13]. In fact, we argue that the latency of the view-invariant body posture representation observed here is indeed modulated by top-down prediction. That is, in case of pure bottom-up processing along the visual hierarchy, one would expect a high-level representation such as view-invariant body posture in LOTC and aIPL to have a larger processing latency compared to low-level representations – which is not what we observed. Instead, the dRSA peak latency for view-invariant body posture in LOTC (~70 msec) is shorter than the peak latency for the view-dependent body posture and pixelwise models (~200 and ~110 msec, resp.; Fig. 2a and 3a), suggesting that the observed neural representation of view-invariant body posture is at least partly shifted ahead in time by top-down prediction. This interpretation is in line with the observation that prediction shifts the neural representation of the position of simple (apparently) moving objects closer to real-time[14,17], and could explain the subjective experience of perceiving our dynamic environment in real-time rather than lagged[17,37]. While this interpretation might seem intuitive regarding our conscious experience of the world, a real-time position representation fails to explain how we are able to act promptly (e.g., catch a ball), as we would need some representation of the stimulus trajectory to reach the motor cortex clearly well ahead of real-time. Albeit speculative, it may therefore be that the close-to-real-time body posture representation serves our conscious experience, while in contrast the clearly predictive motion representations serve prompt behavior. Additionally, the exact prediction latencies as identified here, as well as those observed in previous studies, likely depend on stimulus and task characteristics. Future work is needed to identify which factors (e.g., speed, predictability, prior knowledge, attention) determine prediction latencies, and to clarify the apparent difference in representational latency between body posture (i.e., close to real-time) and body motion (i.e., clearly predictive).

Temporal spread of neural representations

Interestingly, most neural representations were not exclusive to an exact latency, but spread over time, as quantified 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). Such spread could be caused by representations themselves being temporally sustained, or by variability in predictive latency across timepoints within a stimulus, across stimuli, or across subjects. One would expect more spread for higher-level predictions (i.e., of features changing at a longer timescale), and more precise timing (i.e., sharper peak) for low-level models with predictions close in time to the actual sensory input. This is indeed what we observe (Fig. 2a and 3c), i.e., while view-invariant body motion shows most representational spread on both sides of the peak, this is less so for view-dependent body motion. If in fact representational spread partly reflects a sustained representation, this might be explained by gradually, rather than abruptly, emerging predictions.

For lagged representations, this could indicate that representations were not immediately overwritten by new incoming visual input but remained partly activated. The fact that lagged 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[38] 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 post-hoc ANOVA with factors model and ROI ($F = 13.3$, $p = 1.5e\text{-}3$). 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[39].

The temporal forecast window of the brain

Taken together, dynamic RSA allows quantifying the temporal forecast window of the brain. As such, it represents a methodological advancement for studying when our brain represents and predicts the dynamics of the world. It reveals both the peak latency and spread of neural representations of naturalistic dynamic input and does so at several hierarchical levels of processing. The current results go beyond recent evidence for prediction of simple (apparently) moving stimuli in early brain regions[16,17,40,41], by showing that 1) dRSA reveals prediction at both low and high levels of abstraction in complex naturalistic stimuli, and 2) dRSA enables distinguishing these different levels in the same stimulus, thus revealing multiple feature-dependent temporal forecast windows in the brain. In the current dataset it reveals hierarchical predictive representations of the future motion of an observed action, and an absence of post-stimulus motion representations at higher levels at which sensory input is possibly explained away by accurate prediction. While both these observations support predictive processing/coding theories[18–20], future research using dRSA is needed to exclude related theories such as adaptive resonance[33], or Bayesian inference without predictive coding[34].

In principle, dRSA can be applied to any (naturalistic and controlled) dynamic stimulus (e.g., film, soundscape, music, language, motor planning/execution, social interaction) and any biosignal with high temporal resolution (e.g., M/EEG, EcoG, animal electrophysiology, EMG, eye-tracking). 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[42]. 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[43], or used RSA to distinguish words prior to their presentation during naturalistic reading[44]. 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.

Methods

Subjects

Twenty-two healthy human volunteers (mean age, 30 ± 7 years) 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. As we did not have any hypotheses pertaining to differences in sex or gender, we tried to recruit a balanced dataset, in which 13 out of 22 participants identified themselves as female. Sex or gender information was not collected otherwise. We aimed for our findings to be generalizable to the whole population, and therefore combined all subjects irrespective of sex/gender in the group-level statistical analyses. No sex- or gender-based analyses were performed. 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 1 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 counterclockwise. The rational for this 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.

Table 1. Ballet figures per stimulus.

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

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. 4a) that 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).

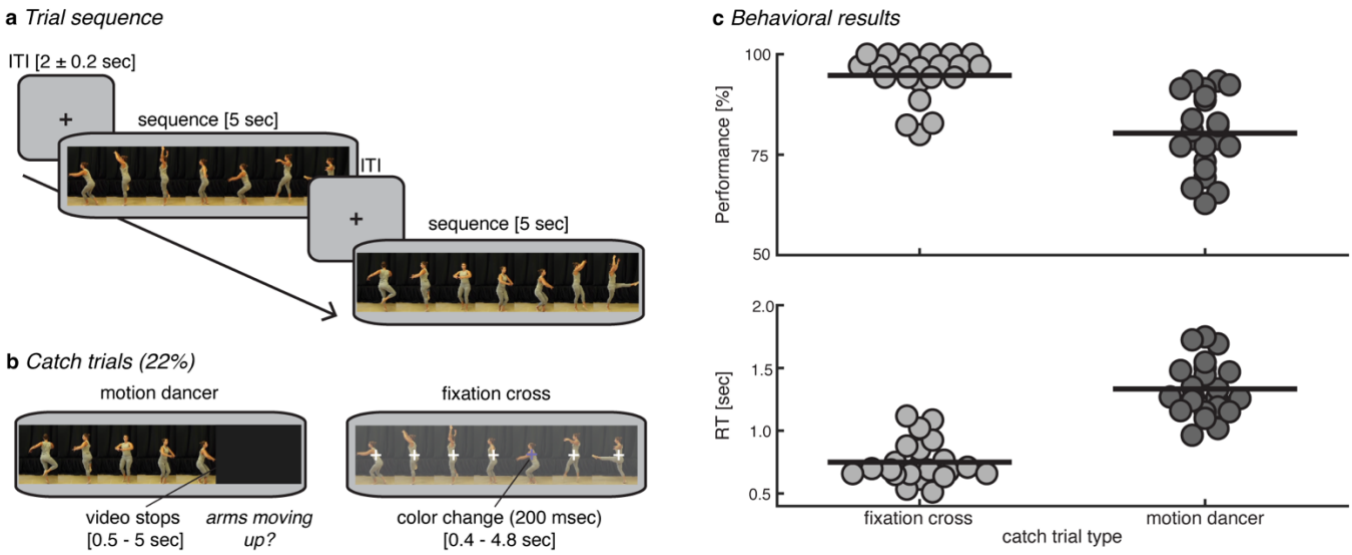


Figure 4. Experimental design and behavioral results. a Trial sequence. A run consisted of 90 sequences of 5 seconds alternated by blank screens with a fixation cross presented for 1.8 – 2.2 seconds (uniform distribution). **b** To ensure attention to the dancer, 15 trials per run were randomly interrupted 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.

To ensure attention to the dance sequences, 20 trials in each run of 90 trials were two types of catch trials (Fig. 4b). 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. 4c). 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. With so many repetitions, one might expect stimulus familiarization throughout the experiment to improve prediction. However, in a post hoc dRSA analysis we did not observe robust differences between the first and last 1/3 of the experiment (see *"Supplementary Discussion"* and Supplementary Figure 6).

Stimulus models

We selected a total of 9 stimulus-feature models described in detail below (Fig. 5), plus a tenth subject-specific model capturing the eye position. The reason for selecting these stimulus-feature models was to capture a comprehensive characterization of the dance sequences that encompasses several hierarchical levels of complexity/abstraction from low-level visual information (pixelwise) up to higher-level, perceptually more invariant information (view-invariant body posture and motion). 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.

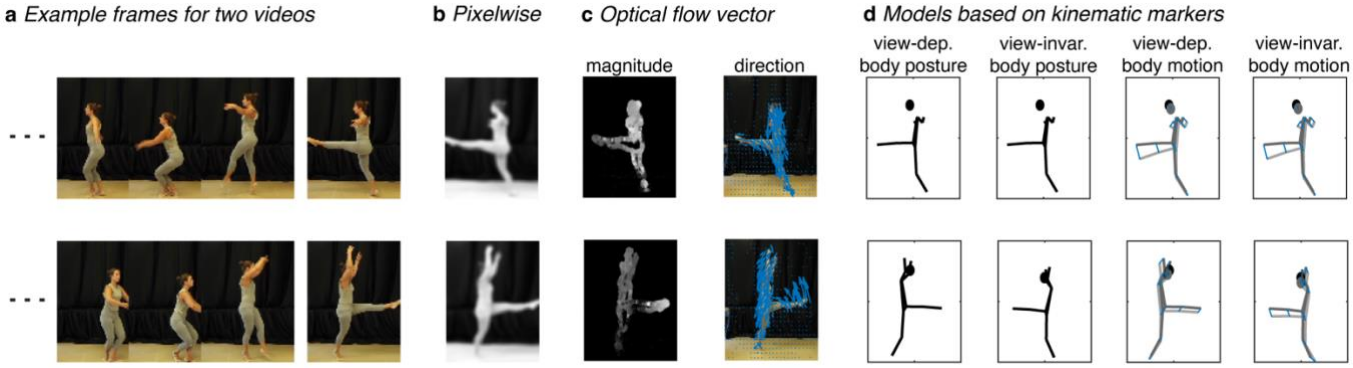


Figure 5. Illustration of stimulus models. a Example frames of two videos (rows) to be correlated for creating the model RDMs. From left to right; **b** the pixelwise smoothed grayscale of the video frame, **c** the magnitude of the optical flow vectors (brighter color is higher magnitude) and the direction 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.

Pixelwise (Fig. 5b): 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. 5c): We converted videos to grayscale and used MATLAB's `estimateFlow.m` implementation of the Farnebäck algorithm[45] 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[46], 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. 5d): 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) $\neq$ dissimilarity(frame1aligned,frame2)), we computed dissimilarities after alignment in both directions, and subsequently averaged over them.

View-dependent body motion (Fig. 5d): 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. 5d): 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[47], 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 (a total of 1.0% of trials was missing; min–max across subjects = 0–5.9%), and the remaining missing samples were interpolated (e.g., due to blinks; a total of 0.3% of samples; min–max across subjects = 0–2.1%). 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. 8b). 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. 8a), 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[48]. 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[49], Fieldtrip[50] and CoSMoMVPA[51] 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[52]) 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[53]), 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[54]. 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[55], 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[56]), 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; Fig. 6) 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)). Besides the purely visual areas, we selected the LOTC, aIPL and PMv because they form the action observation network (AON[57]) and were expected to be important for processing and predicting these specific stimuli[26,58]. 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[51], on the centred data, using 1-Pearson correlation as dissimilarity measure.

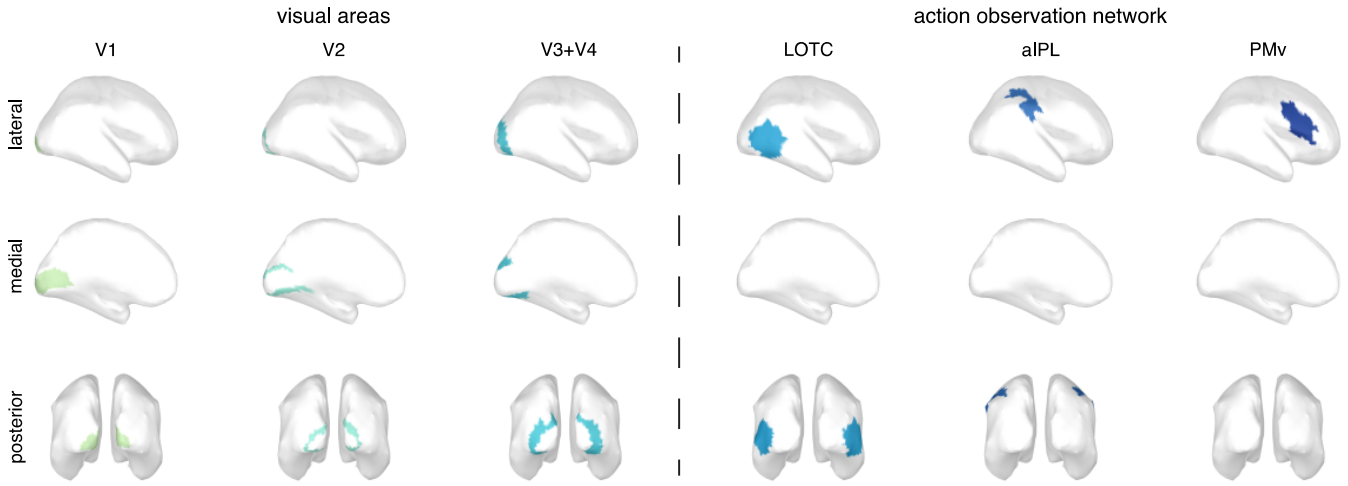


Figure 6. ROI definitions. Cortical regions of interest for main ROI analysis (Fig. 2), based on combinations of parcels as defined according to the Human Connectome Project (HCP) atlas[56]. Here ROIs were a priori defined as follows (ROI name = atlas parcels): V1 = V1 [487 vertices]; V2 = V2 [491 vertices]; 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.

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 SNR, neural and model RDMs were temporally smoothed using a 30 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 below for simulated data (see *"Simulations"*, and particularly Fig. 8a). For completeness, the dRSA results using simple correlations are illustrated in Supplementary Figure 2. While these results (Supplementary Fig. 2) do reflect true shared variance between neural and model RDMs, caution is warranted with their interpretation as it is difficult to separate the contribution of different model RDMs. Note that both shared variance between models and temporal autocorrelation within models are not a problem with the dRSA approach per se, but rather inherent to naturalistic (dynamic) stimuli. We argue that our followed approach explained in the next paragraph offers a solution to separate the contribution of different model RDMs.

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. 8a). 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 preventing 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. A large difference in scale between models could cause early PCA components to simply pick up the model with the largest scale, which arbitrarily depends on dissimilarity measure (i.e., Euclidean distance used for the eye-tracker RDM has a much larger scale than corr-1 used for the other model RDMs). 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. 8b, c and d).

Temporal subsampling

Our dRSA pipeline involved one additional important step to increase the generalizability of the results as well as the SNR. Imagine that at a given time point t_x, in all 14 videos the ballet dancer is in the middle of a figure and moving relatively slowly and predictably. At t_x prediction is possible far in advance, resulting in an early predictive dRSA peak. In contrast, if at t_y there is a transition between ballet figures in all 14 videos, prediction is not possible very far in advance, resulting in a later (or no) predictive dRSA peak. Last, if at t_z stimulus predictability is different for each of the 14 videos (likely the most occurring scenario), the predictive peak reflects an average thereof. In other words, stimulus-specific feature trajectories influence how far in advance a stimulus can be predicted, and thus the structure of the neural and model RDMs depends on the exact arbitrary pairwise alignment of the 14 ballet dancing sequences (and the alignment of event boundaries at different timescales[59], such as onsets and offsets of individual ballet figures, subfigures, smaller motions, etc.). Consequently, the pattern across stimulus time in the 2-dimensional dRSA matrix is idiosyncratic and heterogeneous. If videos would have started a little earlier or later into the first ballet figure, the structure of RDMs, and therefore of the dRSA matrix, would look completely different at all

time points. We are not interested in these specific 14 stimuli and their pairwise temporal alignment per se, but rather in dynamic prediction more generally. Ideally one would have an infinite number of stimuli, i.e., infinitely large RDMs.

To mimic an infinite number of stimuli, and thus minimize the idiosyncrasy of our dRSA results and increase SNR, we applied a temporal subsampling and realignment approach. Specifically, across 1000 iterations, a 3-sec segment was randomly extracted from each of the 14 5-sec stimuli (Fig. 7a; orange boxes). Next, these 14 'new' 3-sec segments were realigned to each other, and RDMs were computed at each realigned time point t_R (Fig. 7b and c). Last, similarity between neural and model RDMs was computed for each realigned neural by model time, resulting in the 2-dimensional dRSA matrix. Crucially, within a single iteration, while a different random 3-sec window was selected for each of the 14 stimuli, for a given stimulus the temporal alignment between neural and model data remained intact (Fig. 7a; dotted vertical orange lines). This subsampling approach accomplishes a different pairwise alignment of the 14 stimuli on each iteration, thus creating completely new RDMs and temporal structure in the dRSA results. This is illustrated in Figure 7d: First, the dashed black line illustrates the results for V1 without the temporal subsampling step (taking the whole 5-sec videos; see Supplementary Fig. 3 for the complete results). Second, the light gray lines illustrate the dRSA results on an example of 10 iterations. There is quite some variability in the shape of the curves, as well as the peak timings, which dependent on the arbitrary pairwise stimulus alignment on that given iteration. Last, the full black lines indicate the dRSA results after averaging the 1000 iterations. 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*'). To summarize, this approach mimics having many more video stimuli that are differently aligned to each other, increasing SNR, and making the results become more generalizable (at least to other sets of ballet dancing videos).

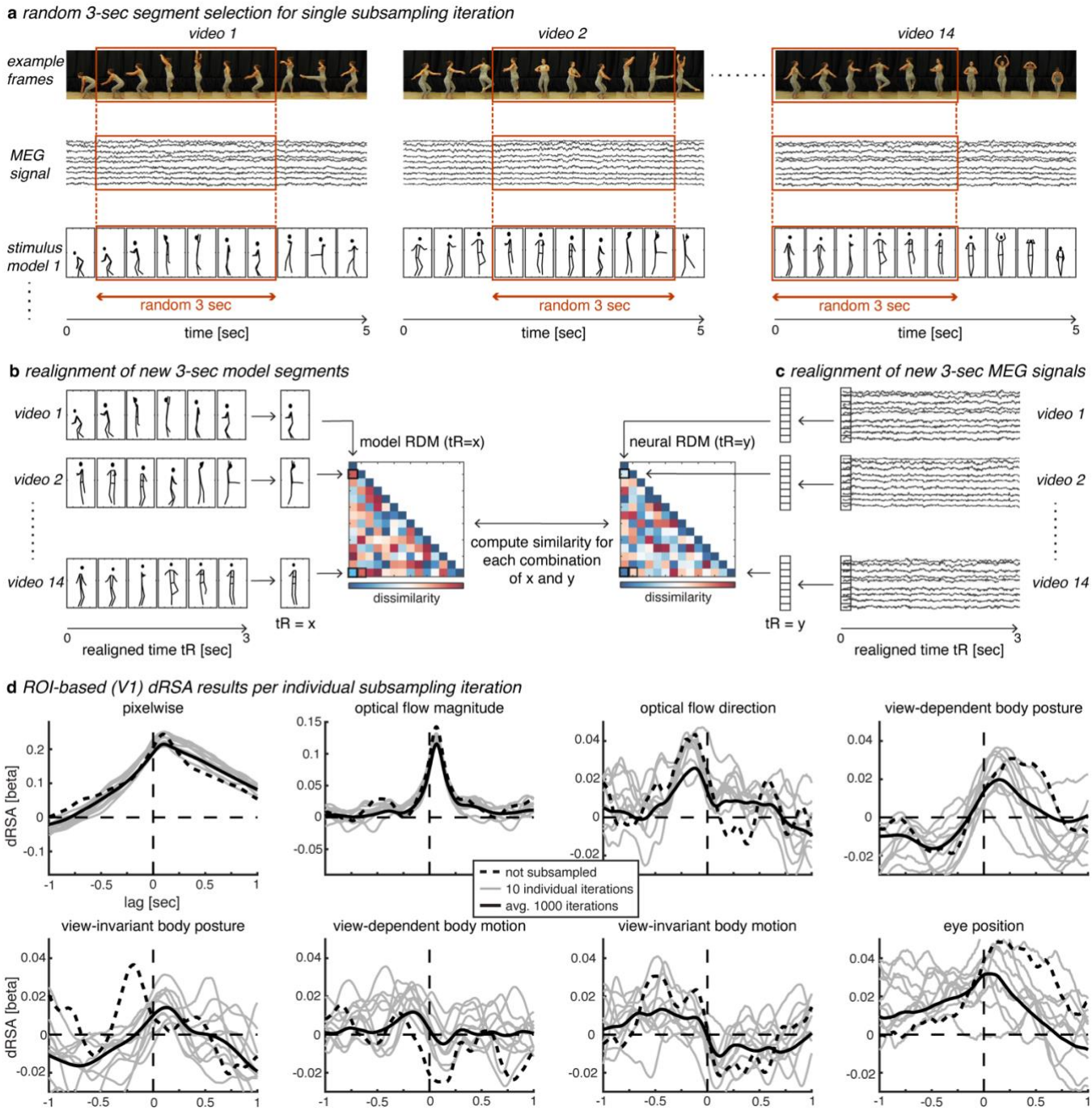


Figure 7. Temporal subsampling and realignment. a On each subsampling iteration, a 3-sec segment is randomly selected independently for each of the 14 stimuli (orange boxes). For a given stimulus, the same 3-sec window is selected for both neural and model data, thus keeping temporal alignment between those intact (indicated by vertical orange dotted lines). **b** and **c** Next, the 14 'new' 3-sec segments are realigned, after which model and neural RDMs are computed at each realigned time point t_R. Last, similarity is computed for each combination of realigned model time (x) and neural time (y). Note that the steps in **a, b** and **c**, as well as the last step to compute the dRSA plot, are all done within a single subsampling iteration. After 1000 subsampling iterations, dRSA results are averaged over iterations. **d** ROI-based dRSA results for V1. The dashed black line indicates dRSA results for the exact same analysis as presented in Figure 2a, but without the temporal subsampling and realignment steps, i.e., dRSA analysis was performed directly on the original 5-sec stimuli (see Supplementary Fig. 3 for other ROIs and statistical significance). The gray lines indicate single 3-sec subsampling iterations, while the full black line indicates the average over 1000 subsampling iterations (equal to V1 in Fig. 2a).

Note that 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, sufficiently large such that results are generalizable rather than specific to the stimulus set (which we solved here with the temporal subsampling and realignment procedure), and dynamic enough that the

structure of the RDM at different time points is distinguishable (i.e., stimulus pairs need to be 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[60]. 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[61], 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. In the main results (Fig. 2a), lenient significant intervals are indicated by thick horizontal bars with colors matching the respective ROI line plots, while strict significant intervals are indicated with thin black or white horizontal bars inside the colored bars. Only significant clusters larger than what can be expected by chance survive this procedure.

Simulations

The goal of the simulations was to find a dRSA pipeline that was effective at 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 most effective at achieving this goal. Note that to prevent double-dipping[62], we first fine-tuned the dRSA pipeline described above on the simulated data, before subsequently applying 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. 8a, see also Supplementary Figure 5 for shared variance, i.e., correlation squared). 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. 8b), 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. 8b), 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).

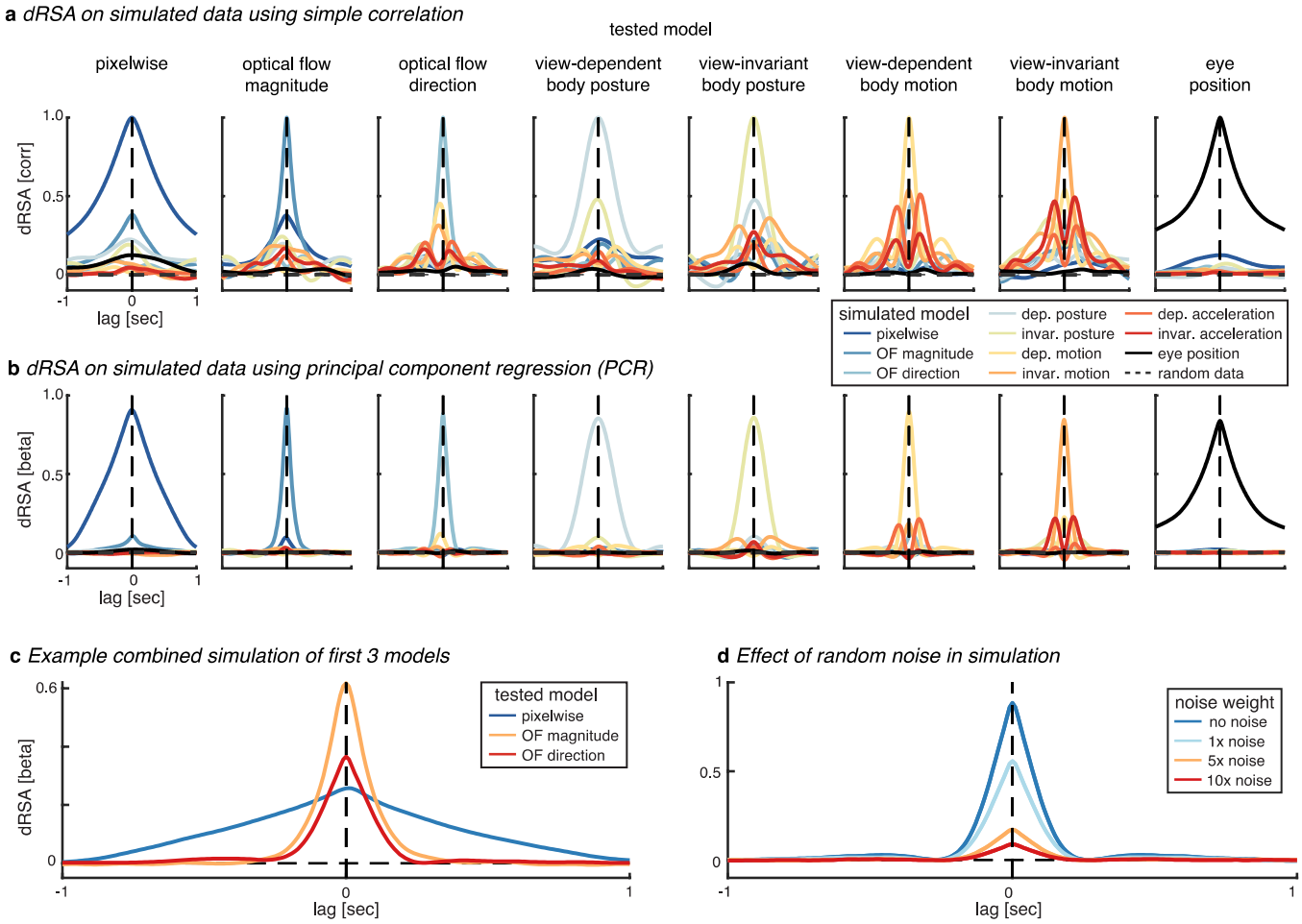


Figure 8. Simulations. a Correlation as similarity measure for dRSA. Colors indicate separate neural simulations of individual model RDMs. Columns indicate which model was tested. These results illustrate the effect of shared variance between various models and the effect of temporal autocorrelation 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.

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. 8c). 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. 8d). 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).

Additionally, using the simulated data we explored several, perhaps more tractable, alternative approaches such as partial correlation to deal with the large amount of (covarying) predictor RDMs. Two example alternatives are illustrated in Supplementary Figure 4. For the first approach based on partial correlation, peak latencies of the cross-model correlations were extracted from Figure 8a, after which a partial correlation that included the model of interest at a certain timepoint t_x, as well as all other models at their respective peak latencies in the cross-correlation with the model of interest (Supplementary Fig. 4b). Since some cross-model correlations displayed two peaks (e.g., view-invariant body posture and motion; see Fig. 8a), a maximum of 2 peak latencies were selected for each to-partial-out model. The second approach based on partial correlation included besides the model of interest all other models after down sampling (250 msec; Supplementary Fig. 4c). The simulations illustrated in Supplementary Fig. 4 indicate that both alternatives perform worse at attenuating effects of co-varying models, and in some cases, they produce quite extreme (spurious) negative values, likely due to multicollinearity. Crucially, applying these approaches to the real data resulted in similar patterns as PCR (data not shown), indicating that the here presented results are not specific to the PCR approach.

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[35]. 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[36]. 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 an ANOVA with factors model (2) and ROI (4), where we included 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 a similar ANOVA with factors model (3) and ROI (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. 8a), which also affects the shape of the

dRSA curve resulting from the PCR method (Fig. 8b). 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. 8b, thus capturing the pure effect of the model autocorrelation in terms of temporal information spread in the dRSA curve. Note that there is another source of autocorrelation, i.e., in the neural data. However, one needs to remember that the neural RDM is computed by first averaging ~36 trials of the exact same video, before computing pairwise dissimilarity between all 14 stimuli. By averaging over ~36 trials any neural autocorrelation that is not stimulus related is canceled out. Any remaining neural autocorrelation is thus related to the stimuli, which is exactly what we are interested in. That is, if this stimulus-related neural autocorrelation is stronger than can be explained by the model autocorrelation, then it must be that the neural signal contains a spread-out representation of the model.

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[63], 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.

Data availability

As the authors are still using this MEG dataset for a successive project, it is currently only available upon reasonable request. As soon as the authors are finished with the dataset, it will be made publicly available. However, note that a complete dataset of a single subject is already available at OSF such that the analysis pipeline can be tested (DOI: 10.17605/OSF.IO/ZK42F). Additionally, the single-subject 'Source Data' for all main results (Fig. 2a and b, Fig. 3a and c, and Supplementary Fig. 1) are provided with this paper, and available on the OSF at DOI 10.17605/OSF.IO/ZK42F.

Code availability

All analysis scripts, the experiment script, and the stimuli used for the here presented results are freely available at https://github.com/Ingmar-de-Vries/DynamicPredictions.

References

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. *The Journal of Neuroscience* **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. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cogn Sci* **22**, 764–779 (2018).
12. 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).
13. Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F. & Seth, A. K. Expectations accelerate entry of visual stimuli into awareness. *J Vis* **15**, (2015).
14. 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).
15. Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W. & Muckli, L. Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience* **30**, 2960–2966 (2010).
16. 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).

17. Johnson, P. A. *et al.* Position representations of moving objects align with real-time position in the early visual response. *Elife* **12**, (2023).
18. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **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. 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).
22. Gayet, S. & Peelen, M. V. Preparatory attention incorporates contextual expectations. *Current Biology* **32**, 687-692.e6 (2022).
23. 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).
24. Kok, P. & De Lange, F. P. Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Current Biology* **24**, 1531–1535 (2014).
25. 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).
26. Kilner, J. M., Friston, K. J. & Frith, C. D. Predictive coding: An account of the mirror neuron system. *Cogn Process* **8**, 159–166 (2007).
27. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **36**, 181–204 (2013).
28. 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).
29. Summerfield, C. & De Lange, F. P. Expectation in perceptual decision making: Neural and computational mechanisms. *Nat Rev Neurosci* **15**, 745–756 (2014).
30. Friston, K. The free-energy principle: A unified brain theory? *Nat Rev Neurosci* **11**, 127–138 (2010).
31. Walsh, K. S., McGovern, D. P., Clark, A. & O'Connell, R. G. Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences* vol. 1464 242–268 Preprint at https://doi.org/10.1111/nyas.14321 (2020).
32. Van Essen, D. C. & Maunsell, J. H. R. Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci* **6**, 370–375 (1983).
33. Lee, T. S. & Mumford, D. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A* **20**, 1434 (2003).

34. Aitchison, L. & Lengyel, M. With or without you: predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology* vol. 46 219–227 Preprint at https://doi.org/10.1016/j.conb.2017.08.010 (2017).
35. 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).
36. Smulders, F. T. Y. Simplifying jackknifing of ERPs and getting more out of it: Retrieving estimates of participants' latencies. *Psychophysiology* **47**, 387–392 (2010).
37. Hogendoorn, H. Perception in real-time: predicting the present, reconstructing the past. *Trends Cogn Sci* **26**, 128–141 (2022).
38. Baumgarten, T. J. *et al.* Neural integration underlying naturalistic prediction flexibly adapts to varying sensory input rate. *Nat Commun* **12**, 1–14 (2021).
39. Merigan, W. H. & Maunsell, J. H. R. How Parallel are the Primate Visual Pathways? *Annu Rev Neurosci* **16**, 369–402 (1993).
40. 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).
41. 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.
42. King, J. R. & Dehaene, S. Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn Sci* **18**, 203–210 (2014).
43. 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).
44. Wang, L. *et al.* Neural evidence for the prediction of animacy features during language comprehension: Evidence from MEG and EEG representational similarity analysis. *Journal of Neuroscience* **40**, 3278–3291 (2020).
45. 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).
46. 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.
47. 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).
48. 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).

49. 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).
50. 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).
51. 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).
52. Taulu, S. & Simola, J. Spatiotemporal Signal Space Separation Method for Rejecting Nearby Interference in MEG Measurements. *Phys Med Biol* **51**, 1759 (2006).
53. Guggenmos, M., Sterzer, P. & Cichy, R. M. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *Neuroimage* **173**, 434–447 (2018).
54. Gaser, C. & Dahnke, R. CAT-a computational anatomy toolbox for the analysis of structural MRI data. in *OHBM* 336–348 (2016).
55. Hämäläinen, M. S. & Ilmoniemi, R. J. Interpreting magnetic fields of the brain: minimum norm estimates. *Med Biol Eng Comput* **32**, 35–42 (1994).
56. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
57. Wurm, M. F. & Lingnau, A. Decoding actions at different levels of abstraction. *Journal of Neuroscience* **35**, 7727–7735 (2015).
58. 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).
59. Baldassano, C. *et al.* Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709-721.e5 (2017).
60. Silver, N. C. & Dunlap, W. P. Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology* **72**, 146–148 (1987).
61. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* **164**, 177–190 (2007).
62. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. & Baker, C. I. Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci* **12**, 535–540 (2009).
63. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral 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

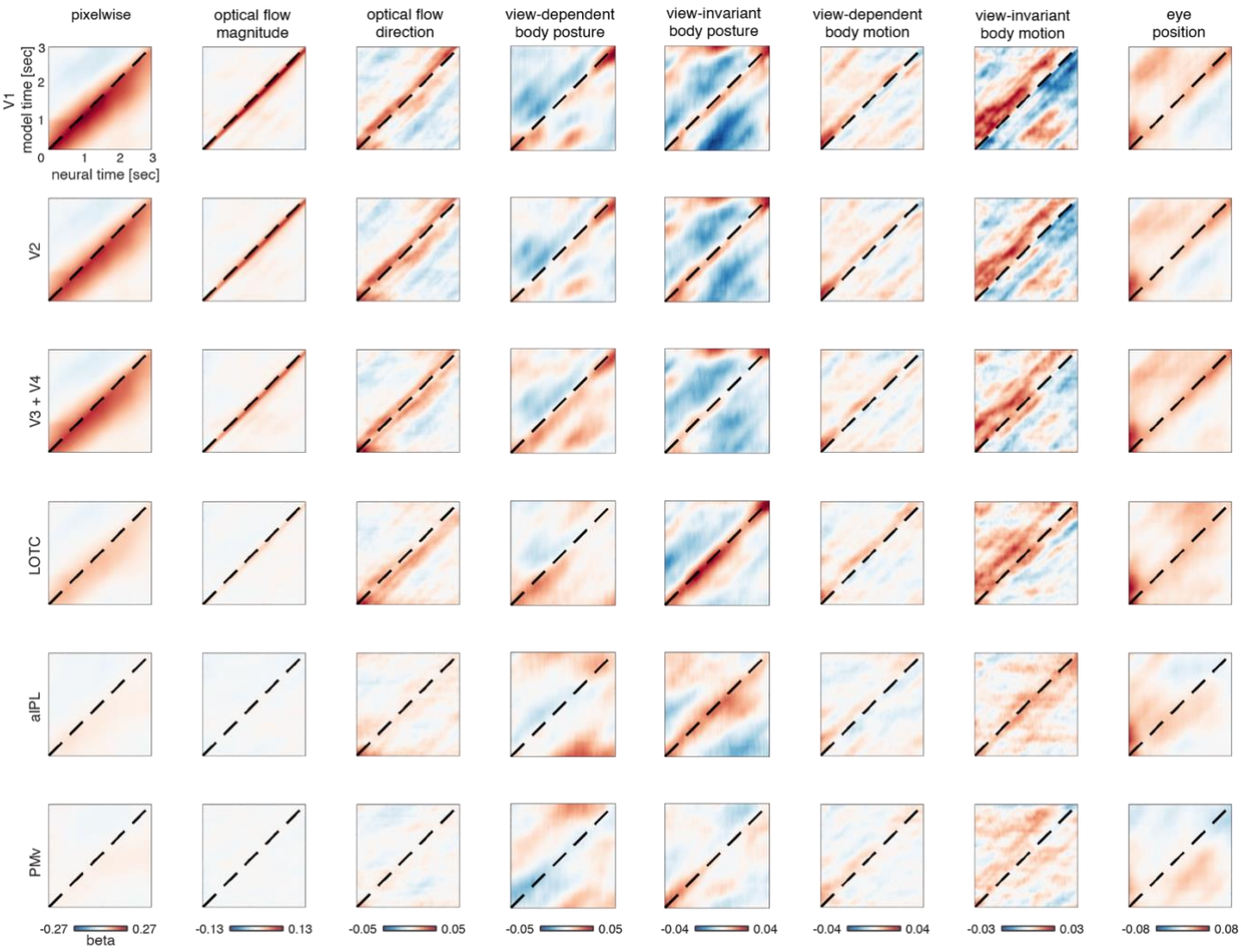
Writing—original draft: IEJV

Writing—review & editing: IEJV, MFW

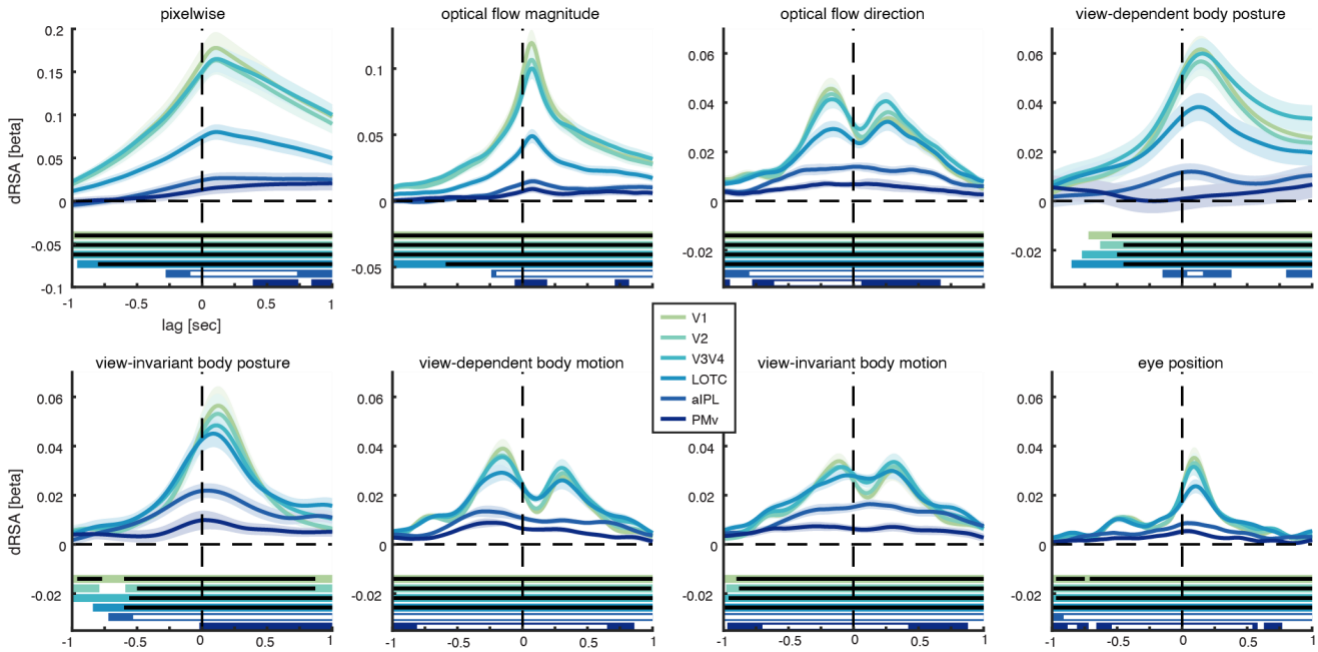
Competing interests

The authors declare that they have no competing interests.

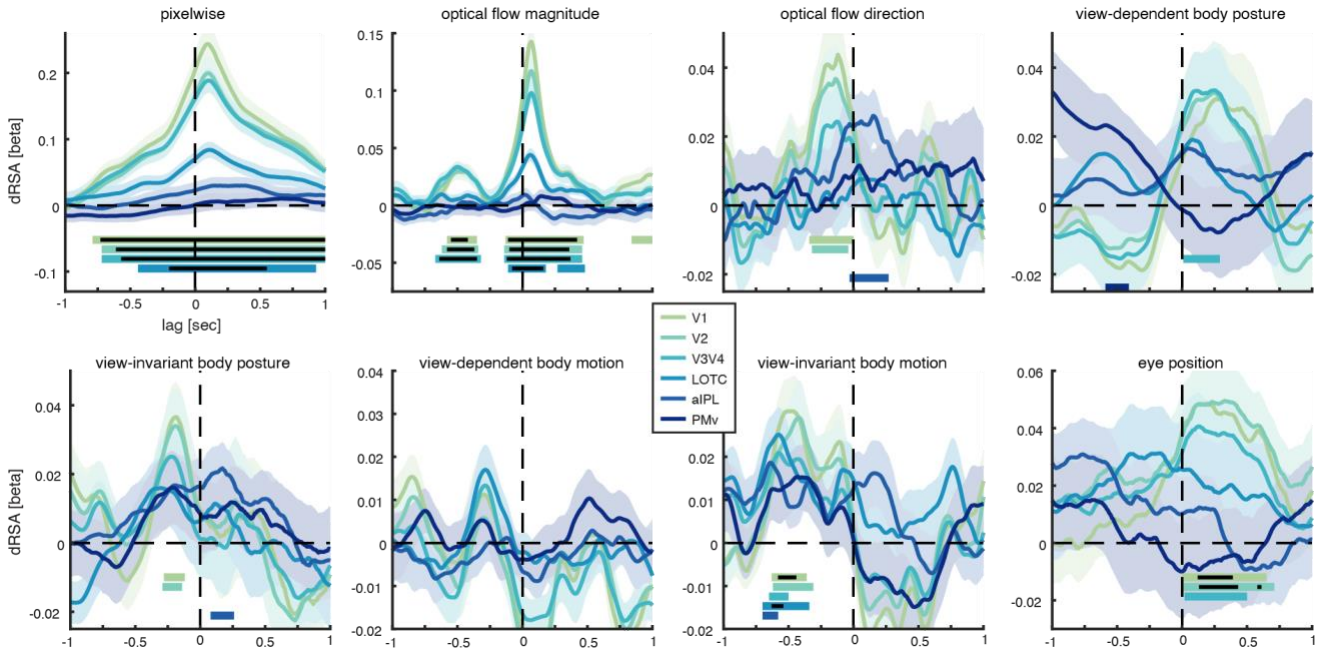
Supplementary Figures



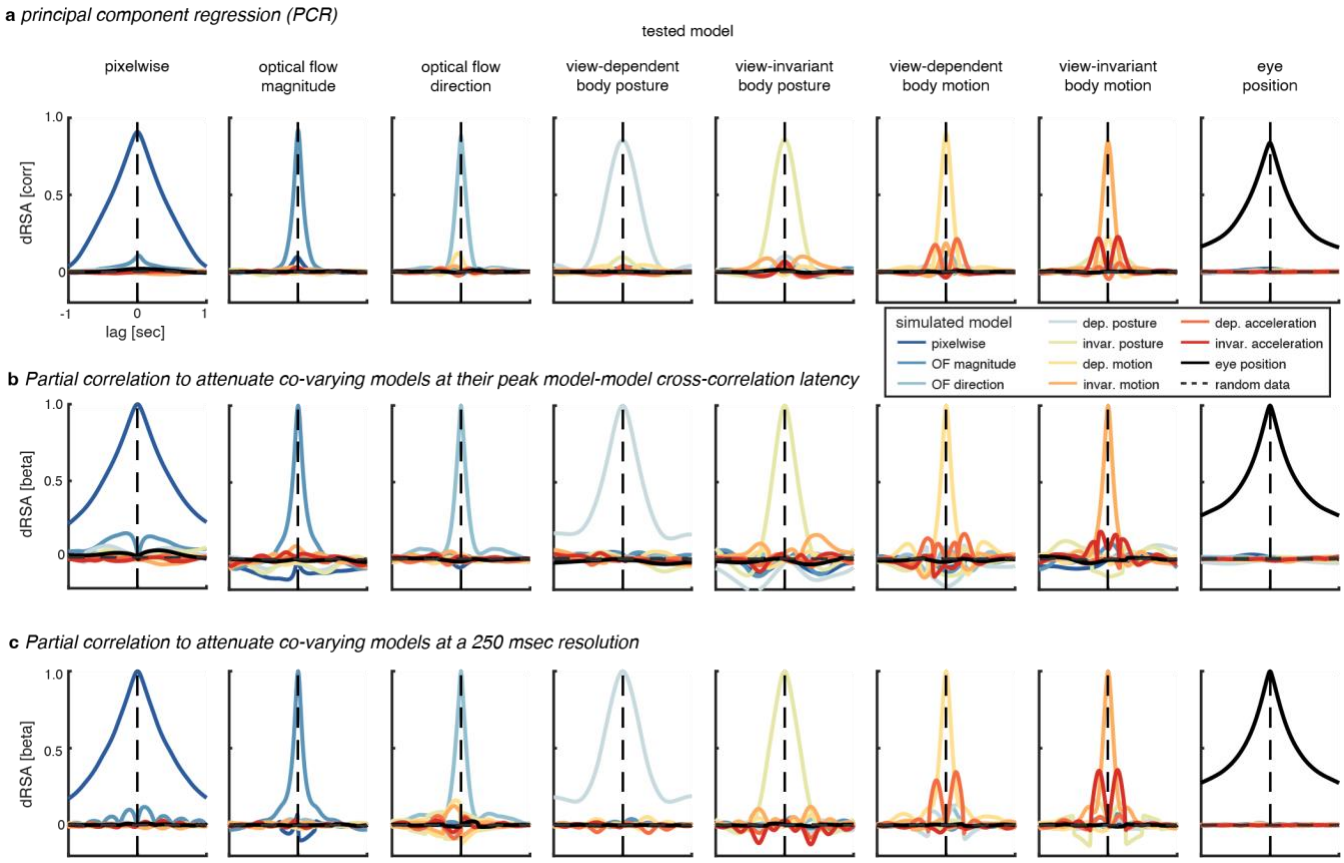
Supplementary Figure 1. Neural-time by model-time two-dimensional dRSA. Dynamic RSA regression weights, illustrated as 2-dimensional model-time by neural-time dRSA matrix, where stronger weights in the lower right triangle indicate that the neural representation is lagging behind a model representation, 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.



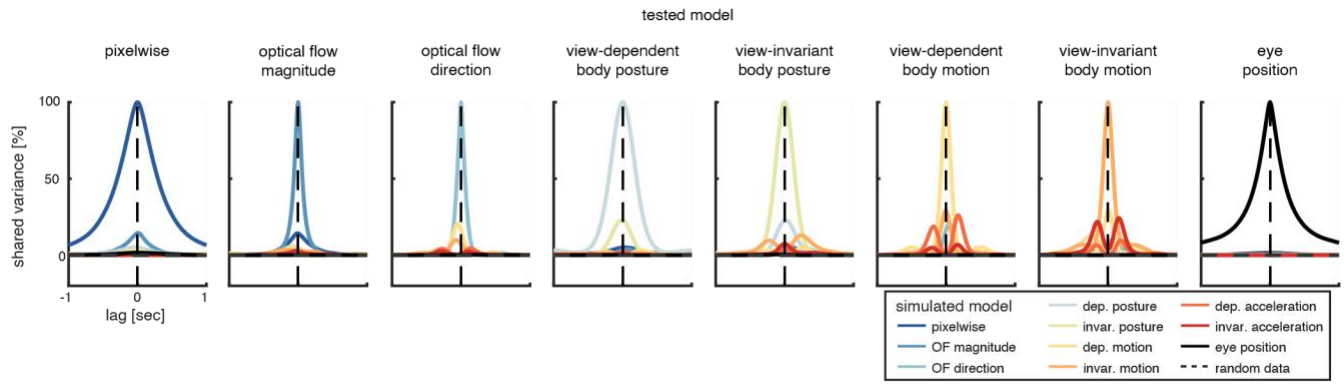
Supplementary Figure 2. ROI-based dRSA results with correlation instead of PCR as similarity measure. Exact same analysis as presented in Figure 2 with the same conventions for colors, subplots, significant intervals, etc., but with correlation coefficient instead of regression weights as similarity measure. That is, the neural RDM is correlated with each model RDM separately.



Supplementary Figure 3. ROI-based dRSA results without temporal subsampling step. Exact same analysis as presented in Figure 2 with the same conventions for colors, subplots, significant intervals, etc., but without the temporal subsampling and realignment step. That is, the dRSA analysis is performed directly on the original 5-sec stimuli.



Supplementary Figure 4. Comparison of similarity measures on simulated data. a Regression weight as similarity measure for dRSA (as in Fig. 8b). Colors indicate separate neural simulations of individual model RDMs. Columns indicate which model was tested. **b** Partial correlation as similarity measure for dRSA which included besides the model of interest all other models at the latency at which each individual model correlated strongest with the model of interest. Specifically, first peak latencies of the cross-model correlations were extracted from Figure 8a, after which a partial correlation that included the model of interest at a certain timepoint t_x, as well as all other models at their respective peak latencies in the cross-correlation with the model of interest. **c** Partial correlation as similarity measure for dRSA that included besides the model of interest all other models after downsampling (250 msec). Importantly, these results indicate that PCR is most effective at extracting only the model of interest from the simulated neural RDM, while attenuating the effect of co-varying models.



Supplementary Figure 5. Shared variance between models. Exact same data as figure 8a (i.e., cross-model correlations) but squared and multiplied by 100%.

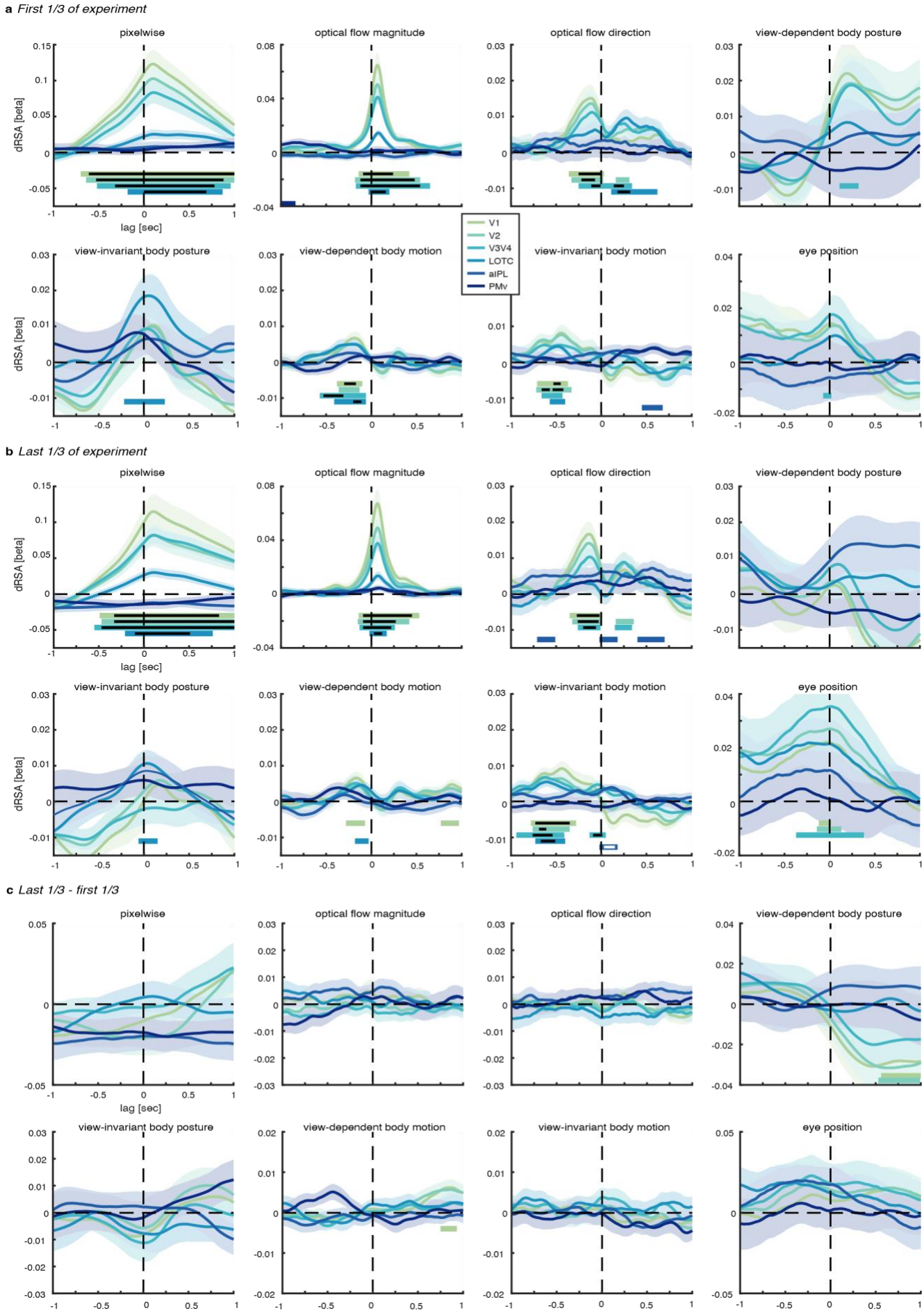
Supplementary Discussion

No effect of stimulus familiarization throughout experiment

Given that the 14 unique videos were each repeated ~36.4 times, one might expect stimulus familiarization throughout the experiment to improve prediction, for instance by lengthening the forecast window, or by strengthening the predictive representation. To test this, we ran the dRSA analysis on the first and last 1/3 of the trials (Supplementary Fig. 6). Most importantly, we did not find significant differences in the predictive

representations. Interestingly, we did observe a reduction in the lagged representation of view-dependent body posture in the later trials, and a trend for a reduction in the lagged representation of optical flow direction in later trials, both of which would be in line with a predictive processing account: i.e., lagged representations would only reflect prediction errors and would thus be expected to reduce with more stimulus knowledge. However, only the effect in view-dependent body motion was significant, and only at our more lenient statistical threshold ($p < 0.05$), thus making this interpretation highly speculative.

One explanation for not finding a robust difference is that it was very difficult for subjects to learn the ballet sequences, as the 14 different combinations of 4 ballet figures, selected from only 5 possible figures, made it hard to predict the next figure at any given time. E.g., if the first figure was a bow, the second could still be an arabesque, passé or jump, and was thus not predictable (see Table 1). This was confirmed by the fact that when anecdotally asked afterwards, subjects did not subjectively experience that they could predict the next ballet figure in the sequence. We therefore believe that the observed predictions are mainly enabled by prior knowledge on biological motion (and gravitation) that is already available to the subjects, rather than new stimulus knowledge/familiarization acquired throughout the experiment. To test the effect of stimulus familiarization (or prior stimulus knowledge) properly, it would be interesting in future research to e.g., compare expert observers to novices, or compare observers pre- and post-training of completely new stimulus material (i.e., not biological motion). Last, note that even if dRSA values are generally lower (due to lower SNR when averaging over less trials), the results are relatively comparable to the main results (Fig. 2), indicating that dRSA works also with less data. In other words, if one so wishes, in future dRSA research it would be possible to use more unique stimuli (i.e., larger RDMs) with fewer repetitions.


a First 1/3 of experiment
b Last 1/3 of experiment
c Last 1/3 - first 1/3
pixelwise
optical flow magnitude
optical flow direction
view-dependent body posture
view-invariant body posture
view-dependent body motion
view-invariant body motion
eye position
dRSA [beta]
lag [sec]
V1
V2
V3V4
LOTC
aIPL
PMv

Supplementary Figure 6. ROI-based dRSA results on first or last 1/3 of trials. Exact same analysis as presented in Figure 2 with the same conventions for colors, subplots, significant intervals, etc., but on only first (**a**) or last (**b**) 1/3 of trials. **c** The difference between results in **a** and **b**.