Rotation-tolerant representations elucidate the time-course of high-level object processing

Denise Moerel[1], Tijl Grootswagers[1,2], Amanda K. Robinson[1,3], Patrick Engeler[1], Alex O. Holcombe[1], Thomas A. Carlson[1]

1. School of Psychology, University of Sydney, Sydney, Australia
2. The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia
3. Queensland Brain Institute, The University of Queensland, Brisbane, Australia

ABSTRACT

Humans have little difficulty recognising visual objects in many circumstances, despite the very different retinal images that result from different viewpoints. One source of variability is 2-D rotation, where an object seen from different perspectives results in different orientations. Here, we studied how the brain transforms rotated object images into object representations that are tolerant to rotation. We measured time-varying electroencephalography responses to objects in eight orientations, presented at either 5 Hz or 20 Hz. We used multivariate classification to assess at what point in time rotation-tolerant object information emerged, and whether we could disrupt the rotation-tolerant object processing by presenting stimuli rapidly (20 Hz) to limit the depth of processing. We compared this to fixed-rotation measures of object decoding, where the classifier is trained and tested on the same orientation. Our results showed that both fixed-rotation and rotation-tolerant object decoding emerged at an early stage of processing, less than 100 ms after stimulus onset. However, rotation-tolerant information peaked later than fixed-rotation information, suggesting rotation-tolerant object representations are most robust during a late stage of processing, around 200 ms after stimulus onset. Both fixed-rotation and rotation-tolerant object information was lower for the 20 Hz compared to 5 Hz presentation rate, which suggests that object information processing is disrupted, but not eliminated, for fast presentation rates. Our results show that object information arises at similar times in the brain regardless of whether it is investigated with the fixed-rotation or rotation-tolerant object decoding method, but it is the later stage of processing that reconciles different viewpoints into a single rotation-tolerant representation.

1. INTRODUCTION

Humans have little difficulty recognising objects across changes in rotation, size, position, and lighting. This implies that the brain creates and maintains representations of objects that are robust to these changes (DiCarlo & Cox, 2007; Rust & DiCarlo, 2010). Robust object representations are advantageous because they provide efficient neural encoding by reducing multiple specific object representations (e.g., different viewpoints) into one. This suggests that a given image elicits image-specific neural responses based on low-level features and is then transformed into a general object representation. Previous work has used multivariate decoding methods to provide insight into the dynamics of visual object representations in the human brain (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019). However, these studies have not assessed whether the decoded object information from the brain is driven by image-specific low-level features or whether it represents an object representation that is invariant, or tolerant, to transformations such as rotation. Here, we use the term 'tolerant' (DiCarlo et al., 2012; Rust & DiCarlo, 2010; Zoccolan et al., 2007) to highlight the dynamic nature of this representation, whereby the visual system initially codes low-level image-specific features, but successive higher-level processes form object representations that are increasingly tolerant to transformations such as rotation. Comparing image-specific and transformation-tolerant representations can yield insight into how objects are represented in the human brain and the process by which this specific to general transformation occurs.

Non-human primate studies have shown that object representations that are tolerant to different transformations emerge in the ventral visual stream. For instance, single neurons in the inferior temporal cortex (IT) are tolerant to rotation (Ratan Murty & Arun, 2015) as well as size and position (Ito et al., 1995; Hung et al., 2005; Zoccolan et al., 2007; Li et al., 2009) and image contrast (Zoccolan et al., 2007), whereas neurons in earlier visual areas such as primary visual cortex respond to image specific features (Hubel & Wiesel, 1963). In addition,

tolerance to image transformations increases as information propagates along the ventral visual stream from V4 to IT (Rust & DiCarlo, 2010). A recent non-human primate study found that size-tolerance developed early after stimulus onset, followed by position-tolerant signals, and then by rotation-tolerant and view-tolerant signals (Ratan Murty & Arun, 2017). For clarity, rotation-tolerant refers to rotations in the picture plane, which do not reveal new object features, whereas view-tolerance refers to rotations that have a depth component, such that different aspects of the object become (in)visible. Together, these studies from non-human primates provide evidence for the emergence of transformation-tolerant object representations as information propagates along the ventral visual stream.

In humans, functional magnetic resonance imaging (fMRI) studies have used the repetition suppression paradigm to investigate the role of different areas along the ventral visual stream in view-tolerant object recognition (Andresen et al., 2009; Grill-Spector et al., 1999; James et al., 2002; Vuilleumier et al., 2002). The findings were mixed, with evidence for view-tolerance in the ventral temporo-occipital cortex (James et al., 2002), left fusiform cortex (Vuilleumier et al., 2002) and left fusiform/occipito-temporal sulcus (Andresen et al., 2009), but little evidence for view-tolerance in the right fusiform cortex (Vuilleumier et al., 2002). Another study did not find evidence for view-tolerance in the lateral occipital complex (Grill-Spector et al., 1999), which is located in earlier in the ventral visual stream, but this was contradicted by a study that used multivariate decoding methods with fMRI (Eger et al., 2008). These fMRI studies suggest higher level regions within the ventral stream are important for transformation-tolerant object recognition. However, they do not provide insight into the time-course of this process.

Previous decoding work on magnetoencephalography (MEG) and electroencephalography (EEG) data has investigated the time-course of the emergence of object information in the brain (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019). Specifically, the studies looked at three different

levels of categorical abstraction: the “animacy level” (judgments of whether an image represented something that can move on its own volition), high-level categories (e.g., mammal, fruit, tool, etc.), and the object category level (e.g., giraffe, apple, hammer, etc.). The results showed evidence for object-specific information in the brain earlier than 100 ms after stimulus onset in each of these studies. However, a linear classifier will use all information that helps to distinguish two objects from the neural signal, so the decoded object information is likely to be driven both by low-level differences between objects as well as higher-level object representations. This means that decoding times found in previous work might not be a good reflection of when the brain has truly established a tolerant or invariant representation of the object.

Studies that have investigated the emergence of high-level category information have found a different temporal signature compared to object exemplar decoding (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019). In these studies, the first decoding peak for object exemplar decoding is around 120 ms, but this peak is not usually observed in high level category decoding. Instead, the high-level category decoding peak is found around 200 ms after stimulus onset. Information about animacy peaks later, usually around 400 ms after stimulus onset. This suggests that at 200 ms, a representation of the high-level object category has emerged and information from a specific exemplar is generalised to other exemplars within that category. This time-course is also consistent with object decoding when the training and test set contain different exemplars of the same object (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019), again suggesting that the information has generalised to other exemplars within the category. It is possible that the time-course of the emergence of rotation-tolerant object representations resembles that of high-level category decoding, as both access representations that are more tolerant to low-level pixel-wise differences between stimulus images.

A few studies have used time-resolved neuroimaging to provide insight into how robust object representations emerge throughout the human visual system. Using MEG, Carlson and colleagues (2011) showed position-tolerant object representations emerged around 100 ms after stimulus onset. Isik and colleagues (2014) used MEG to investigate the emergence of position- and size-tolerant information in object processing and found that tolerant representations emerged later than non-tolerant information. Size-tolerant information peaked first, with a later peak for position-tolerant information. In addition, tolerance to size and position increased along the ventral stream. Supporting this succession of invariant representations, EEG evidence showed tolerance to viewpoint followed size- and position-tolerance (Karimi-Rouzbahani et al., 2017). These findings from humans are in line with the findings in non-human primates reviewed above (Ratan Murty & Arun, 2017). Together, the findings from non-human primates and humans suggest that transformation-tolerant representations emerge while undergoing additional processing stages within the ventral visual stream.

The studies described above have investigated the tolerance to different transformations but have not explored the tolerance to rotation specifically. It is therefore still unclear how *rotation-tolerant* object representations emerge in the human brain. Rotation is arguably a more complex transformation than the scaling and translation needed to reconcile size- and position-invariance, which may be why rotation invariance emerges later in non-human primates (Ratan Murty & Arun, 2017). Behavioural studies in humans have shown that rotation can disrupt the recognition of objects to some extent, depending on whether the object is familiar or not (Edelman & Bülthoff, 1992; Jolicoeur & Milliken, 1989; Lawson & Humphreys, 1996). However, rotation generally does not significantly influence the ability to categorise objects. For instance, performance on the categorisation of human and animal images was affected very little by inversion (Rousselet et al., 2003) and rotation (Guyonneau et al., 2006), suggesting that category-level representations are robust to rotation.

To investigate how the visual system reconciles different rotations into an invariant representation, we investigated the time-course of rotation-tolerant information processing in the human brain in this study. In addition, we asked whether we could disrupt the rotation-tolerant object processing by presenting stimuli at faster speeds, as previous work has shown that faster presentation rates limit visual information processing (Collins et al., 2018; Grootswagers, Robinson, & Carlson, 2019; Grootswagers, Robinson, Shatek, et al., 2019; McKeeff et al., 2007; Robinson et al., 2019), particularly high-level information processing (Holcombe, 2009; McKeeff et al., 2007). We presented sequences of object stimuli in eight in-plane rotations, at 5 Hz and at 20 Hz, and used multivariate decoding methods to investigate the coding of rotation-tolerant information over time. To ensure the classifier could not rely on differences in the retinotopic pattern between stimulus images, and was forced to generalise between different rotations, the classifier was trained and tested on different rotations of the same object. We contrasted this with the object information obtained through fixed-rotation object decoding methods (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019), where the classifier was not trained to generalise across different object orientations. We use decoding peak times to compare time-courses between conditions, as the optimal decoding time has been shown to correlate with object categorisation behaviour (Ritchie et al., 2015). The comparison between rotation-tolerant and fixed-rotation object decoding can show how the brain transforms image-specific representations into a general object representation.

The results for the 5 Hz presentation condition showed that object information emerged around a similar time for fixed-rotation and rotation-tolerant object decoding analysis. However, the rotation-tolerant information peaked later compared to the object information found with the fixed-rotation test. Adding controls for the contribution of low-level processes did not substantially change the time-course of the rotation-tolerant object decoding. That is, the representational similarity analysis showed a similar time-course for rotation-tolerant

object compared to the rotation-tolerant object decoding. Taken together, the results show that fixed-rotation object representations reflect low-level object features to some degree, yielding the early decoding peaks of approximately 120 ms. When focussing on higher level object representations by investigating rotation-tolerant object processing, a later peak time was observed around 200 ms. This suggests that at the later stage of processing different viewpoints are reconciled into a more robust object representation, leading to recognition. These findings highlight the importance of focussing on the coding of rotation-tolerant object information to investigate the time-course of object representations in the brain.

2. METHODS

Some aspects of this study were preregistered via Open Science Framework: https://osf.io/r93es. The experiment code, analysis codes, results, and figures can also be found there. The raw and pre-processed data for this study can be found on OpenNeuro: https://doi.org/10.18112/openneuro.ds004252.v1.0.1.

2.1. Participants

Sixteen healthy adults participated in the study (age range = 19-60 years, 14 female / 2 male, 14 right-handed / 2 left-handed). All participants reported having normal or corrected to normal vision. Participants received monetary payment or undergraduate course credit for their participation. The study was approved by the ethics committee of The University of Sydney, and participants provided informed consent.

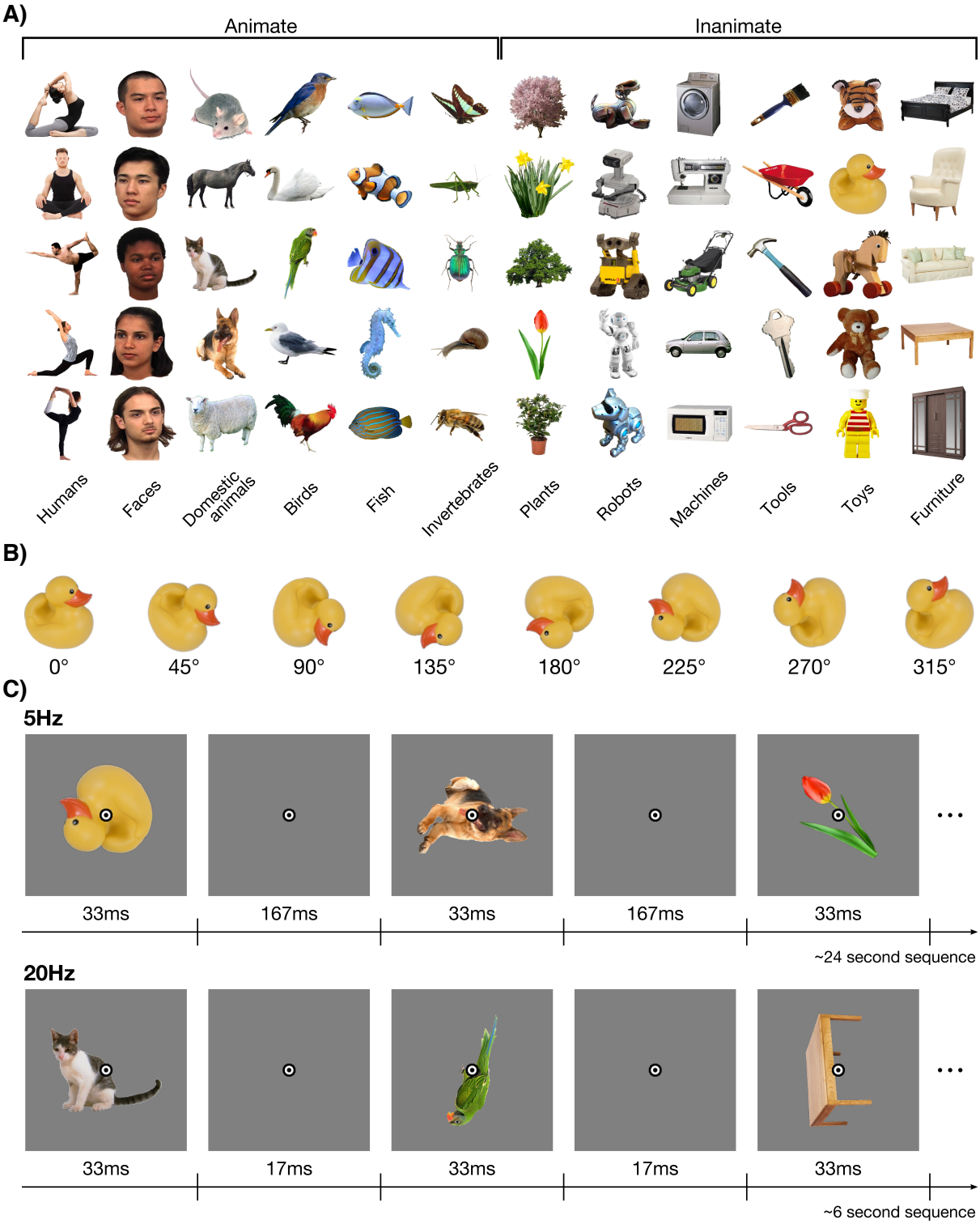

Figure 1. Stimuli and experimental design. A) The 60 stimuli were grouped into two categorical levels: 1) animacy, and 2) object category. Each object category contained five exemplars. The 0° rotation is displayed in A. **B)** The eight rotations that each stimulus was presented in ranged from 0° to 315° in steps of 45°. **C)** The stimulus sequences were presented at two frequencies: 5 Hz and 20 Hz. The stimuli were presented for 33 ms in both presentation frequency conditions, so the interval between two consecutive stimuli was 167 ms for the 5 Hz condition and 17 ms for the 20 Hz condition.

2.2. Stimuli and experiment procedure

Figure 1A provides an overview of the stimuli. The stimulus set consisted of 60 objects, taken from Contini et al. (2020) and an online image bank (www.pngimg.com). The images can be grouped at the animacy level, with 30 animate and 30 inanimate images, or at the category level, with 12 categories (e.g., humans, birds, toys). Each category contained five objects. We presented the object stimuli in eight rotations: 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° (see Figure 1B). The stimuli were presented at fixation on a mid-grey background (RGB = 128,128,128), at a distance of approximately 60 cm. The stimuli were 512 x 512 pixels (approximately 12.50 degrees visual angle) in size.

Participants viewed 120 sequences of images, and each sequence contained 120 object images. There were 15 repetitions per object, rotation, and presentation frequency condition, resulting in a total of 14,400 stimulus presentations. In each sequence the 60 images were shown twice in two different rotations. The stimulus items in each sequence were presented in random order, and each object-rotation combination in the sequence was unique compared to the other stimuli in the sequence. Four consecutive blocks contained a single repeat of all objects and rotations. We presented the image sequences at two frequencies, 5 Hz and 20 Hz. Figure 1C shows part of an example sequence for both presentation frequency conditions. Stimuli were presented for 33 ms in both presentation frequency conditions. In the 5 Hz condition, the interval between successive stimuli was 167 ms, whereas the interval was 17 ms in the 20 Hz condition. Each sequence had a single stimulus presentation frequency, and the counterbalanced presentation frequencies were presented in pseudo-random order. The experiment took approximately 45 minutes to complete.

Participants were instructed to maintain fixation at a central bullseye for the duration of the experiment and to respond to an infrequent target by pressing a button. The purpose of this task was to keep participants engaged in the experiment, and to ensure participants

remained fixating at the centre of the objects. The bullseye consisted of two black rings and a white ring in between (see Figure 1C). The target consisted of a black filled circle the same size as the fixation bullseye and was presented instead of the fixation bullseye for the duration of a stimulus presentation (33 ms). Each sequence contained between two and four targets. We placed the targets at random in each sequence, using the following constraints. Two targets were spaced at least 10 stimulus presentations apart, and the first and last 10 stimulus presentations were never a target. We counted responses within 1 s of the target as correct. Participants had an average accuracy of 89% on the fixation colour change detection task (SEM = 2.1%, range = 70.5% to 99.7%).

2.3. EEG acquisition and pre-processing

We used a BrainVision ActiChamp system to record continuous EEG data from 128 electrodes, digitised at a sample rate of 1000Hz. The electrodes were placed according to the international standard 10-20 electrode placement system (Oostenveld & Praamstra, 2001), and were referenced online to FCz. We used the EEGlab Matlab toolbox to pre-process the data (Delorme & Makeig, 2004), following a pre-processing pipeline used in previous work (Grootswagers et al., 2021; Robinson et al., 2021). We interpolated bad channels that measured more than 5 standard deviations away from the average, using the kurtosis measure. We re-referenced the data to an average reference and filtered the data using 0.1 Hz high pass and 100 Hz low pass Hamming windowed FIR filters. We down-sampled the data to 250 Hz and created epochs for each stimulus in the sequence from -100 ms to 800 ms relative to stimulus onset.

2.4. Decoding analysis

We applied decoding analyses to determine the time-course of fixed-rotation measures of object information and rotation-tolerant object information, using the CoSMoMVPA toolbox for

Matlab (Oosterhof et al., 2016). We performed the analyses within-subject, separately for the 5 Hz and 20 Hz sequences. For each time-point in the epoch, we trained a linear discriminant analysis (LDA) classifier to distinguish between the 60 objects (see Figure 1A). As we used a 60-way object classification analysis, chance performance was 1.67%. To decode the fixed-rotation object information, we followed the method used in previous work (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019), by training the classifier on the same rotation as used in the test set. For each unique combination of object and rotation, there were 15 repetitions of the identical stimulus over the entire experiment for each presentation speed condition. The training set consisted of 14 of the 15 identical stimulus repetitions of each of the 60 objects and the test set consisted of the left-out stimulus repetition of each object in the same rotation, for example training on 14 repetitions of all 60 objects at 0° rotation and testing on one repetition of all 60 objects at 0°. We repeated this analysis 15 times, leaving out a different stimulus repetition of each of the 60 objects each time, and averaged across decoding accuracies. We also repeated this analysis for all the rotations, and averaged decoding accuracies across them. This analysis was done for each time-point. To decode rotation-tolerant object information, we trained the classifier to distinguish between the 60 objects on 7 different rotations of an object and tested the classifier on the eighth rotation. We ensured that the classifiers were trained and tested on the same amount of data as the fixed-rotation method by always training on 14 repetitions of each object and testing on one repetition of each object, for example training on 14 repetitions of all 60 objects, with 2 repeats per orientation (45°, 90°, 135°, 180°, 225°, 270°, and 315°) and testing on the left-out repetition of all 60 objects at 0°. Note that we tested the classifier on the exact same trials for the fixed-rotation and the rotation-tolerant object decoding analyses. The only difference between the different analyses was that the models were always trained and tested on the same rotation angle in the fixed-rotation object decoding analysis, consistent with previous studies (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers,

Robinson, & Carlson, 2019), whereas rotation-tolerant models were trained and tested on different rotation angles. It is therefore possible to directly compare classifier accuracies between fixed-rotation and rotation-tolerant object decoding analyses.

2.5. Exploratory analyses

2.5.1. Channel searchlight

To investigate which EEG channels were driving the classification accuracies, we obtained time-varying topographies by performing an exploratory channel-searchlight analysis, following an established pipeline (Grootswagers, Robinson, Shatek, et al., 2019; Robinson et al., 2019). We constructed a local cluster for each EEG channel by taking the closest 4 neighbouring channels and ran the decoding analysis described above separately for the fixed-rotation and rotation-tolerant object coding analyses, for the 5 Hz and 20 Hz conditions. The decoding accuracy for each cluster was stored in the centre channel, resulting in a time-by-channel map of the decoding accuracies for each condition, separately for each participant.

2.5.2. Representational Similarity Analysis

We used the Representational Similarity Analysis (RSA) framework (Kriegeskorte et al., 2008; Kriegeskorte & Kievit, 2013) to investigate the temporal dynamics of rotation-tolerant object representations, while accounting for low-level visual information. This framework allows us to compare the representational structure obtained by EEG to different models of object representations, using representational dissimilarity matrices (RDMs) that represent the dissimilarity between item activations. Figure 2 gives an overview of this analysis. We created a 480 by 480 neural RDM from the EEG data, which coded for the neural dissimilarity between each combination of unique stimuli (60 object images in 8 rotations, see Figure 2B). We used cross-validated pairwise decoding accuracy as a measure of dissimilarity, as a more dissimilar neural response between items results in a higher decoding accuracy. We obtained a neural

RDM for each participant, separately for each time-point and for the 5 Hz and 20 Hz stimulus presentation conditions (Figure 2A). For each time-point, we correlated the lower triangle of the neural RDM to a model RDM that codes for the object in a rotation-tolerant way (Figure 2C). We partialled out five low-level image feature control models from this correlation (Figure 2D). This allows us to obtain the correlation between the EEG activation and the rotation-tolerant object model that is not driven by these low-level image features.

We used the following five low-level image feature control models: a silhouette model, a pixel-wise colour model, a pixel-wise luminance model, a mean luminance model and an RMS contrast model. The silhouette model was based on the binary alpha layer of the vectorised experimental images, using the jaccard distance measure to calculate the dissimilarity between pairs of images. The colour, luminance, and contrast models were based on the CIELAB colour values or luminance values of the images and were not measured on the stimulus presentation screen. Note that although these measures are approximate, the model is based on the (dis)similarity between two images on these measures. The pixel-wise colour model was based on the CIELAB colour values of the vectorised images, using correlation as a distance measure. We used the Matlab rgb2lab function to convert the RGB values of the image to CIELAB. We assumed that the colour space of the screen was sRGB and the white point was D65. The pixel-wise luminance model was based on the greyscale values of the vectorised images, using correlation as a distance measure. In addition, we created two models that were based on a value calculated across the entire image; a mean luminance model and an RMS contrast model (Harrison, 2022). The mean luminance model was based on the mean greyscale values of the stimuli, and the RMS contrast model on the standard deviation of the greyscale values of the stimuli. We used the mean difference as a distance measure. The dissimilarity between the five low-level image feature control models and the rotation-tolerant object model are shown in Figure 2E-F. The three pixel-wise visual

models (silhouette, pixel-wise colour, and pixel-wise luminance) were similar, as were the two models based on the entire image (mean luminance and RMS contrast).

2.5.3. Temporal generalisation

We applied the temporal generalisation approach (Carlson et al., 2011; King & Dehaene, 2014; Meyers et al., 2008) to determine 1) whether object representations at one time-point generalised to other times, and 2) whether the object representations were similar for the 5 Hz and 20 Hz presentation conditions. Instead of training and testing the classifier on a single time-point, the classifier is trained and tested on all combinations of time-points. This results in a training time by test time time-generalisation matrix and allows us to capture similar object representations that occur at different times. To investigate the temporal generalisation between object representations over time, we trained and tested a classifier in the same way as described in the decoding analysis above for each combination of training and testing time-points. We did this separately for the 5 Hz and 20 Hz conditions. To determine the overlap in object representations between the 5 Hz and 20 Hz presentation conditions, we trained a classifier on the 5 Hz condition for a time-point in the epoch and then tested the classifier on every time-point in the 20 Hz condition. We repeated this analysis for all training time-points. In addition, we trained on the 20 Hz condition and tested on the 5 Hz condition, and averaged the two time-generalisation matrices (Grootswagers, Robinson, & Carlson, 2019; Kaiser et al., 2016).

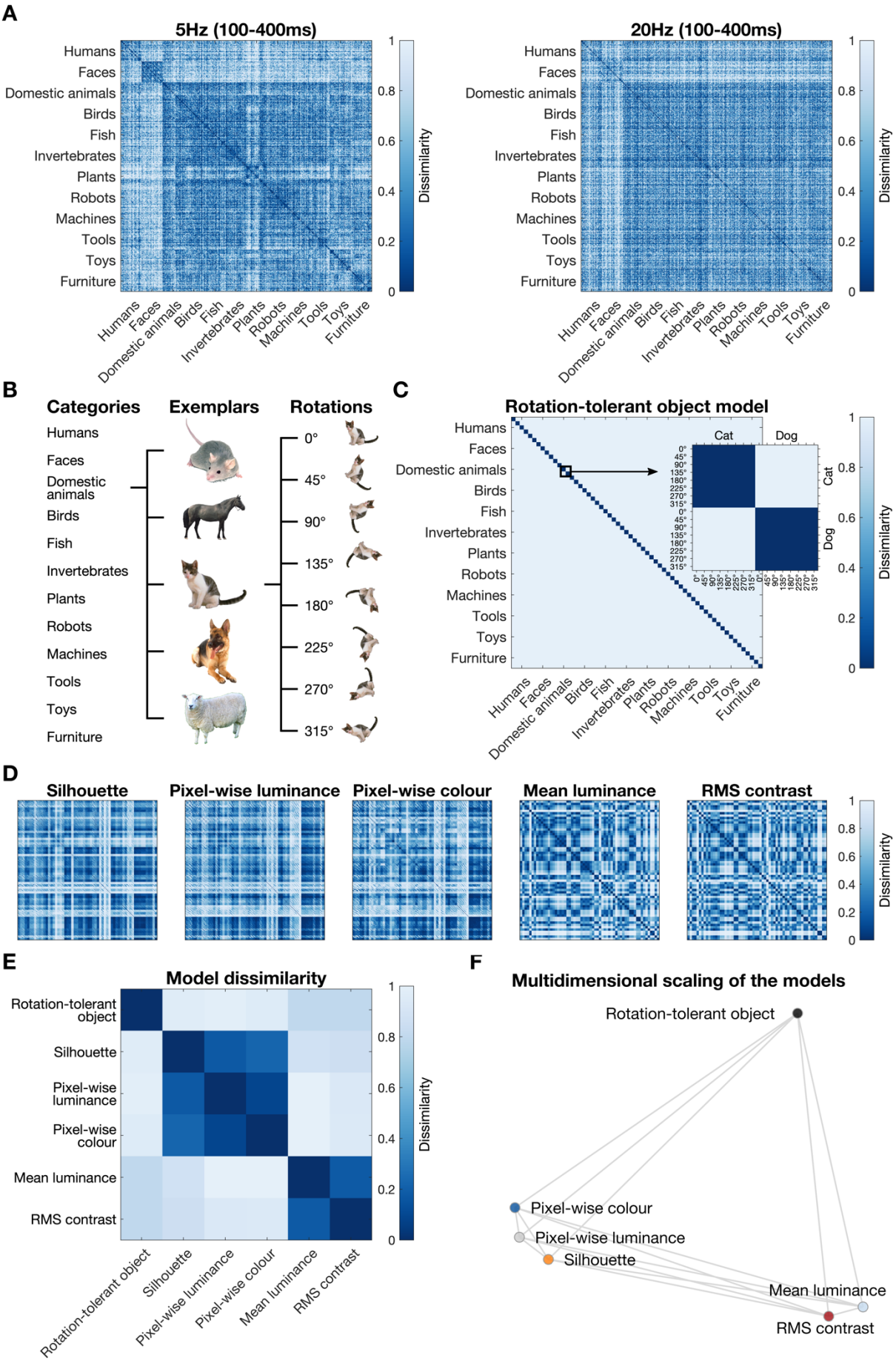
A
5Hz (100-400ms)
20Hz (100-400ms)
Humans
Faces
Domestic animals
Birds
Fish
Invertebrates
Plants
Robots
Machines
Tools
Toys
Furniture
Dissimilarity
0
0.2
0.4
0.6
0.8
1
B
Categories
Exemplars
Rotations
0°
45°
90°
135°
180°
225°
270°
315°
C
Rotation-tolerant object model
Cat
Dog
D
Silhouette
Pixel-wise luminance
Pixel-wise colour
Mean luminance
RMS contrast
E
Model dissimilarity
Rotation-tolerant object
F
Multidimensional scaling of the models
Rotation-tolerant object
Pixel-wise colour
Pixel-wise luminance
Silhouette
Mean luminance
RMS contrast

Figure 2. Models used in the Representational Similarity Analysis. A) The representational dissimilarity matrix (RDM) for the EEG data, averaged across the 100 ms to 400 ms time-window, for the 5 Hz (left) and the 20 Hz (right) condition. Each point in the 480 by 480 matrix represents the dissimilarity between two stimuli. **B)** The labels for all 480 by 480 matrices plotted here: the rows and columns consist of 12 categories x 5 exemplars x 8 rotations. **C)** The rotation-tolerant object model. This model predicts that the response to the same object is very similar, regardless of the orientation in which the object is presented, whereas the response to two different objects is dissimilar. **D)** The low-level image feature control models that were used to control for low-level visual differences between the stimuli. **E)** The dissimilarity between the six different models: the rotation-tolerant object model and the low-level image feature control models. **F)** Shows a projection of the model dissimilarities into 2-dimensional space. We used multi-dimensional scaling to obtain a configuration where the distance between two models approximates the dissimilarity between those models.

2.6. Statistical inference

We used Bayesian statistics (Dienes, 2011; Kass & Raftery, 1995; Morey et al., 2016; Rouder et al., 2009; Wagenmakers, 2007) to determine whether decoding accuracies were at chance level, or instead substantially higher. We used the Bayes Factor R package (Morey & Rouder, 2018) to implement the statistical analyses. We implemented the same analysis for the fixed-rotation and the rotation-tolerant object decoding analyses, for both the 5 Hz and the 20 Hz presentation conditions. For each time-point, we applied a Bayesian t-test with a point prior for the null hypothesis and a notched half-Cauchy prior for the alternative hypothesis to test for directional effects (Morey & Rouder, 2011). The half-Cauchy prior was centred around $d = 0$ (i.e., chance-level decoding accuracy of 1.67%) and had the default width of 0.707 (Jeffreys, 1998; Rouder et al., 2009; Wetzels et al., 2011). We excluded the interval between $d = 0$ and $d = 0.5$ (the “notch”) from the prior (Teichmann et al., 2022). Because this analysis pits a point null hypothesis against a $d > 0.5$ alternative hypothesis, this analysis tests whether the decoding is substantially above chance not just marginally above chance. We repeated this analysis for each time-point. All statistical analyses were applied at the group level.

To test whether there was a difference between the coding of object information found with the fixed-rotation and rotation-tolerant object decoding analyses, we calculated the

difference between these decoding accuracies and used a Bayesian t-test to test the difference score against 0 for each time-point. We used a point prior for the null hypothesis and a full-Cauchy prior for the alternative hypothesis to test for differences in either direction between the coding of object information found with the fixed-rotation and rotation-tolerant object decoding analyses. For consistency with the t-test described above, we used the default prior width of 0.707 with a notch (excluding the interval between d = -0.5 and d = 0.5) from the prior.

We used the same statistical analysis described above for the representational similarity analysis. For the temporal generalisation analysis, we repeated the same statistical analysis for each combination of training and test time-points.

3. RESULTS

3.1. Fixed-rotation and rotation-tolerant object decoding analyses

The temporal dynamics of object representations obtained with fixed-rotation decoding and a rotation-tolerant decoding method are shown in Figure 3. For the 5 Hz presentation condition (Figure 3A), there was clear evidence (BF > 10) for substantially above-chance fixed-rotation object decoding from ~84 ms after stimulus onset. Clear evidence for substantially above chance decoding of the rotation-tolerant object information emerged at a similar time, approximately 92 ms after stimulus onset. However, object coding found through the fixed-rotation method peaked ~116 ms after stimulus onset, whereas the coding of the rotation-tolerant object information peaked at ~192 ms after stimulus onset. The bootstrapped 95% confidence intervals of the peak decoding time did not overlap for the object coding obtained via the two different methods, demonstrating a later peak for rotation-tolerant object information coding. We observed a difference in object coding between the fixed-rotation and rotation-tolerant object decoding methods from ~88 ms after stimulus onset, with stronger fixed-rotation object coding compared to the rotation-tolerant object decoding analysis.

Together, these results suggest that object information emerges around a similar time regardless of whether we trained and tested on the same orientation or on different orientations. However, rotation-tolerant object information peaked later compared to the fixed-rotation analysis, suggesting that rotation-tolerant object representations are most robust at later stages of processing.

We repeated the same analysis for the 20 Hz presentation condition. We aimed to provide insight into the depth of object processing by comparing object coding for fast and slower presentation rates. A recent study showed a decrease in object coding when stimuli are presented at a fast compared to slower presentation rate (Robinson et al., 2019), suggesting that faster presentation rates impaired visual processing. Figure 3B shows the decoding accuracies for object coding for the 20 Hz presentation condition, obtained for the fixed-rotation and rotation-tolerant analyses. Both fixed-rotation and rotation-tolerant object information could be decoded at a similar time. For fixed-rotation decoding, when the classifier was trained and tested on the same rotation, the evidence began to clearly favour substantially above-chance decoding from ~96 ms after stimulus onset, with peak decoding at ~116 ms. There was clear evidence for substantially above-chance rotation-tolerant object decoding from ~112 ms after stimulus onset, peaking at ~120 ms after stimulus onset. There was substantial evidence for stronger object decoding for the fixed-rotation analysis compared to rotation-tolerant analysis starting at ~100 ms after stimulus onset. These results show that rotation-tolerant information still emerges for a high presentation rate, though there was not an obvious late peak as in the slower condition.

Directly comparing the 5 Hz and 20 Hz presentation conditions with a notched Bayesian t-test, object coding was stronger for the 5 Hz condition for both the fixed-rotation analysis, starting at ~92 ms after stimulus onset, and the rotation-tolerant analysis, starting at ~104 ms after stimulus onset. Together, these results suggest that the 20 Hz presentation speed limits, but does not eliminate, both fixed-rotation and rotation-tolerant object processing.

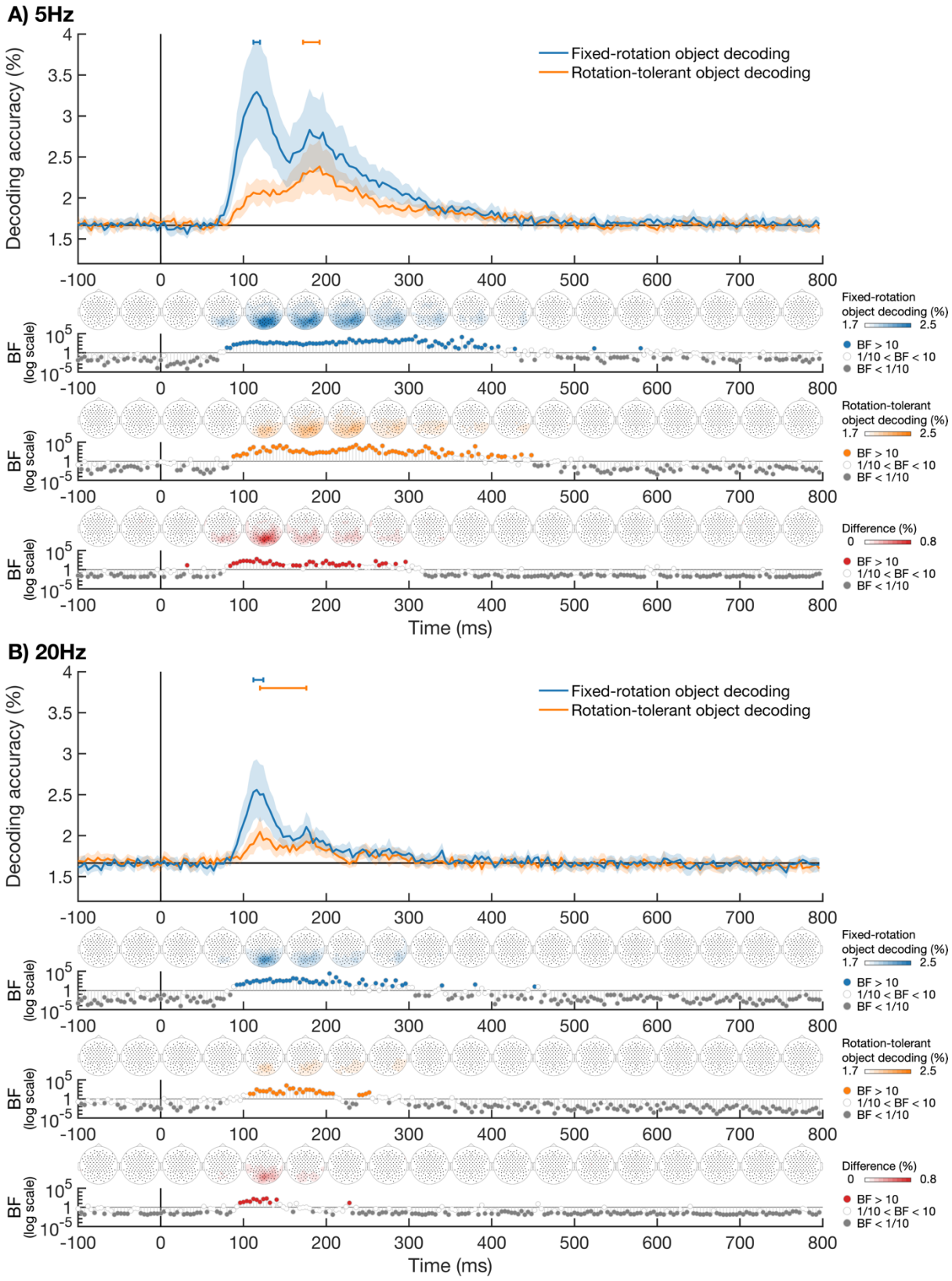

Figure 3. Time-course of object decoding accuracy for 5 Hz (A) and 20 Hz (B) presentation. A) Object decoding data for the 5 Hz presentation. We used linear classifiers to decode object information via the fixed-rotation decoding method (blue), that were trained

and tested on the same rotation, and in a rotation-tolerant way (orange), where we trained and tested on different object rotations. Decoding accuracies are based on 60-way decoding, which means chance is 1.67% decoding accuracy. The shaded areas around the plot lines show the bootstrapped 95% confidence intervals across participants. Bootstrapped 95% confidence intervals of the peak times are displayed as error bars above the peaks. The time-varying topographies, which were obtained from the exploratory channel-searchlight analysis, are displayed below each plot, averaged within 50 ms time bins. We obtained separate topographies for the fixed-rotation (blue) and rotation-tolerant (orange) object decoding methods. Notched Bayes factors (excluding small positive effect sizes from the prior) are shown below the plot. Notched Bayes factors below 1/10 are shown in grey and those above 10 are shown in colour. **B)** Object decoding accuracy for the fixed-rotation analysis (blue) and the rotation-tolerant (orange) analysis for the 20 Hz presentation condition. Plotting conventions are the same as in A.

3.2. Representational similarity analysis

In the analysis described above, we assessed the decoding accuracy for rotation-tolerant object information by training the classifier on all but one rotation and testing the classifier on the left-out rotation. The classifier cannot rely on specific pixel-wise configurations between stimulus images in the rotation-tolerant object decoding analysis and is forced to generalise between different rotations. This decoding analysis is therefore less likely to reflect low-level differences in object features compared to the fixed-rotation object decoding. However, it is possible that the decoding of the rotation-tolerant object is driven in part by low-level visual differences between the object images. We used representational similarity analysis to account for low-level visual differences, by modelling the contribution of different types of low-level differences. This allows us to investigate the time-course of rotation-tolerant object coding when further reducing the contribution of these low-level visual differences.

To investigate whether the visual models can explain variance in the EEG data, we first explore the correlation between the five different low-level image feature models and the EEG data for the 5 Hz and the 20 Hz condition (Figure 4A). The correlation between the EEG data for the 5 Hz condition and the silhouette, pixel-wise luminance, and pixel-wise colour models peaked around 96-100 ms after stimulus onset, suggesting these visual features were

represented in the neural data early in the time-course. The mean luminance and RMS contrast correlation peaked around 120-124 ms after stimulus onset. The pattern of model correlations with the EEG data for the 20 Hz condition was similar, although correlations were lower. There was no correlation with RMS contrast for the 20 Hz condition. Together, these findings indicate that low-level features are most represented relatively early in the neural responses.

Turning to the rotation-tolerant representation, after accounting for the five low-level image feature control models, Figure 4B shows the partial correlation between the EEG data and the rotation-tolerant model for each presentation speed condition. For the 5 Hz presentation condition, substantial evidence for a partial correlation with the rotation-tolerant object representation model was present from ~100 ms after stimulus onset. The peak time of this partial correlation was ~184 ms after stimulus onset. The timings for the 20 Hz condition were very similar, we found substantial evidence for a partial correlation with the rotation-tolerant object model from ~100 ms after stimulus onset, with a peak correlation at ~180 ms after stimulus onset, although for 20 Hz this second peak was not significantly bigger than the first peak. These peak timings for the representational similarity analysis are very similar to those found in the decoding analysis, suggesting rotation-tolerant object representations are evident when images are presented at both 5 Hz and 20 Hz and are unlikely to be driven by low-level visual information about the different objects. Directly comparing the partial correlations of the rotation-tolerant object model with the EEG data from the 5 Hz and 20 Hz conditions, a difference was evident from ~164 ms after stimulus onset, with a higher correlation with the rotation-tolerant object model for the 5 Hz compared to 20 Hz condition. This finding suggests that when the low-level visual information is partialled out, there are no differences in RSA results between the 5 Hz and 20 Hz conditions for the first peak (~120 ms), but there is lower rotation-tolerant information for the second (~200 ms) peak. The early positive correlations between the EEG data and the low-level models noted above were

unsurprising – early visual cortices are known to respond to low-level features. A potentially interesting finding is that for the colour, luminance, and silhouette models, the correlations became *negative* between about 150 and 200 ms. However, a post-hoc notched Bayesian t-tests showed there was no strong evidence for negative correlations. We further explore this finding in the Discussion.

We made a two-dimensional embedding of the RDM to visualise the dynamic representational structure for the images presented at 5 Hz, during the first peak (108 ms) and second peak (184 ms) (Figure 4C). The distance between the images reflects their mean dissimilarity across participants. At 184 ms after stimulus onset, a clear clustering emerges of different rotations of the same images, suggesting the object representation is tolerant to rotation at this time. In addition, the faces and humans become separated from the other object categories at this time, suggesting there is a distinct representation of humans and human faces at 184 ms after stimulus onset.

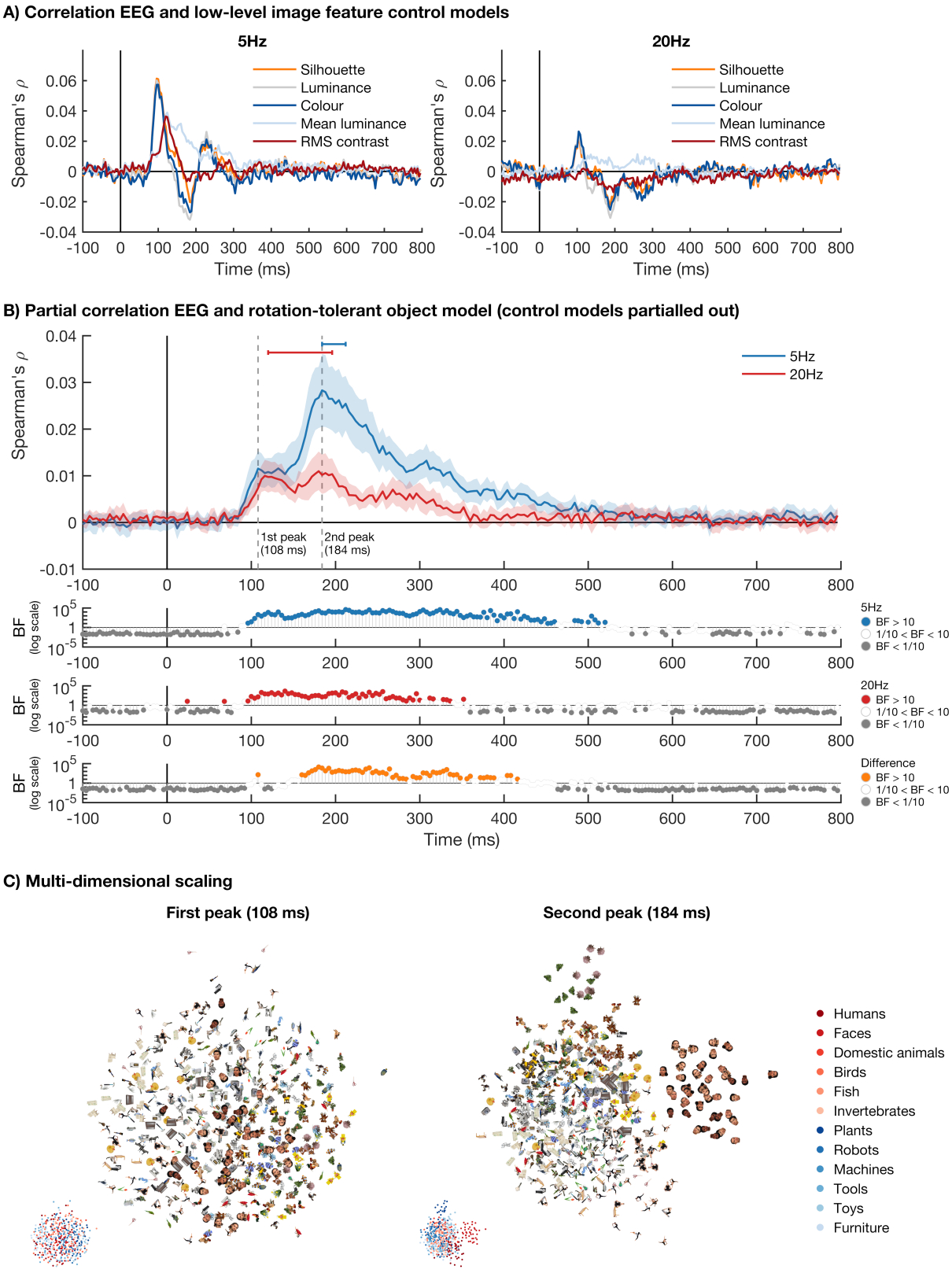

Figure 4. Representation Similarity Analysis results. A) Shows the time-course of the correlations between the five low-level image feature control models and the EEG data for the 5 Hz condition (left panel) and the 20 Hz condition (right panel). **B)** Shows the partial correlation of the rotation-tolerant object model with the EEG data for the 5 Hz condition (blue) and the 20 Hz condition (red) over time, with the five visual control models partialled out. The

bootstrapped 95% confidence intervals across participants are displayed as shaded areas around the plot lines. The error bars above the peaks show the bootstrapped 95% confidence intervals of the peak times. Notched Bayes factors are shown below the plot, with Bayes factors below 1/10 displayed in grey and Bayes factors above 10 shown in the plot colour. The vertical grey dashed lines indicate the first peak (108 ms) and second peak (184 ms) of the partial correlation of the rotation-tolerant object model with the EEG data for the 5 Hz condition. **C)** The neural dissimilarity of the stimuli (presented at 5 Hz) arranged in a two-dimensional space. The distance between two stimuli reflects their pairwise distance, with larger distance between two stimuli reflecting more dissimilar neural responses. The inset on the bottom left corner of each plot shows the same stimuli, with each dot representing a stimulus and the colours representing the 10 different categories. The left panel shows the embedding at 108 ms after stimulus onset, and the right panel at 184 ms after stimulus onset. The categories become noticeably more separated at 184 ms compared to 108 ms, especially the people and faces.

3.3. Temporal generalisation

We used the temporal generalisation approach to investigate 1) whether activation patterns observed at one time-point generalise to other time-points, and 2) whether the object representations were similar between the 5 Hz and 20 Hz presentation conditions. Temporal generalisation results are displayed in Figure 5. To determine whether activation patterns generalise between different time-windows, we trained and tested the classifier to distinguish between objects for all combinations of training and test time-points. We did this separately for the fixed-rotation (Figure 5A) and rotation-tolerant (Figure 5B) object decoding analyses, and for the 5 Hz and 20 Hz presentation conditions. Decoding was strongest along the diagonal for all conditions because the points along the diagonal correspond to testing on the same condition as the classifier was trained on. While some generalisation was present for neighbouring time-points, there was no generalisation between the two different decoding peaks, around 120 ms and 200 ms, which suggests that the decoding in these two time-windows is driven by different patterns of activation. However, there was some evidence for generalisation between the first decoding peak and a later time-window, around 350 ms to 500 ms after stimulus onset.

To determine whether the object representations were similar for the two different presentation rates, we trained on the 5 Hz condition and tested on the 20 Hz condition and vice versa and collapsed the time-generalisation matrices for these two analyses. We transposed the train on 20 Hz, test on 5 Hz time-generalisation matrix before averaging, which means above-chance decoding off the diagonal shows that similar activation occurs at different times for the two presentation speed conditions. We found evidence for cross-decoding between the 5 Hz and 20 Hz conditions for both the fixed-rotation and the rotation-tolerant object decoding analyses. This suggests that the object representations were similar for the two presentation rates. In addition, the decoding was primarily along the diagonal for the cross-decoding between the 5 Hz and 20 Hz conditions, indicating this information was processed around the same times in both speed conditions. However, there was stronger decoding above compared to below the diagonal for the rotation-tolerant object decoding, suggesting there was longer processing for the 5 Hz condition compared to the 20 Hz condition for the rotation-tolerant object coding.

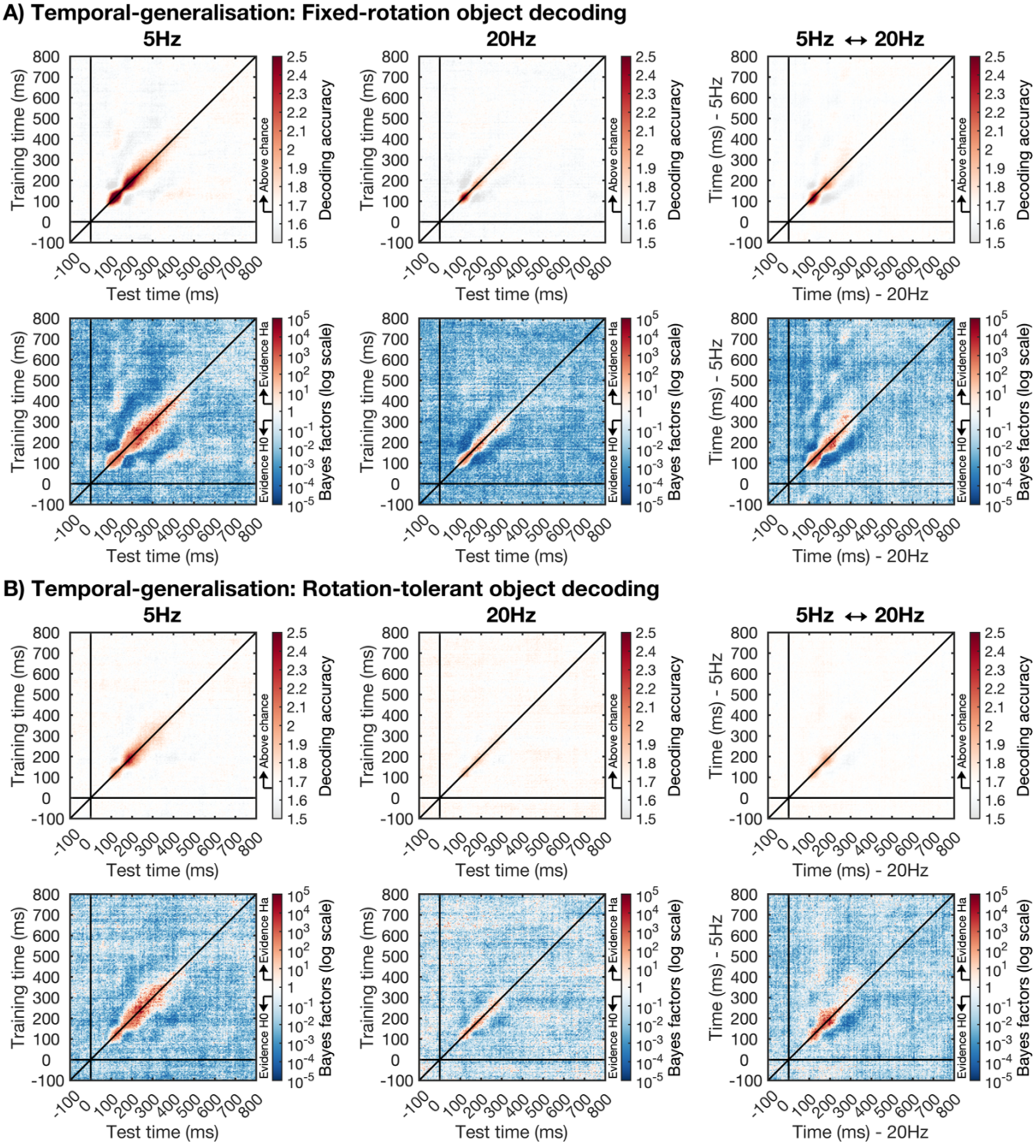

Figure 5. Temporal-generalisation for the fixed-rotation object decoding (A) and rotation-tolerant object decoding (B). We used 60-way object decoding, which means that theoretical chance is at 1.67%. **A)** Shows the temporal-generalisation of decoding accuracies for the fixed-rotation object coding analysis (top row), with the corresponding notched Bayes factors on a logarithmic scale (bottom row). The left panels show the results for the 5 Hz stimulus presentation condition, and the middle panels for the 20 Hz presentation condition. The training times are shown on the y-axis and the test times on the x-axis. The right panels show the cross-decoding between the 5 Hz and 20 Hz condition. For this analysis, we trained on the 5 Hz condition and tested on the 20 Hz condition and vice versa. We averaged the results across the two analyses, transposing the train on 5 Hz test on 20 Hz matrix. The y-axis shows the times for the 5 Hz condition, and the x-axis the times for the 20 Hz condition **B)** shows the temporal-generalisation of decoding accuracies for the rotation-tolerant object

coding (top row) and the corresponding notched Bayes factors on a logarithmic scale (bottom row). Plotting conventions are the same as in A.

4. DISCUSSION

In this study we investigated how the brain transforms images into object representations that are tolerant to rotation. In line with previous studies (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019), our results showed evidence for object coding, including above-chance decoding for the object exemplar when we used the fixed-rotation method of training and testing on the same object rotation. Our results add to these findings by providing evidence for the coding of *rotation-tolerant* object information, and directly contrasting object decoding found through the fixed-rotation decoding method and the rotation-tolerant object decoding method. The evidence began to favour substantially above-chance object decoding around similar times for the fixed-rotation and rotation-tolerant decoding methods, approximately 84-92 ms after stimulus onset. However, the decoding of the rotation-tolerant object information peaked later than the decoding accuracy obtained via the fixed-rotation method, suggesting that rotation-specific information is transformed into a higher-level invariant representation at later stages of processing. When we presented object stimuli at a faster rate (20 Hz), there was still evidence of both fixed-rotation and rotation-tolerant brain representations, although decoding accuracies were lower for stimuli presented at the faster compared to slower (5 Hz) presentation rate. We suspect this happens in part due to masking degrading the signal quality generally, and in part due to the fast presentation speed reducing the depth of processing (Holcombe, 2009; McKeeff et al., 2007) and therefore limiting the amount of rotation-tolerant object information.

Previous work has investigated the time-course of object representations that are invariant to size and position (Isik et al., 2014). The authors found that non-invariant coding peaked at 135 ms, size-invariant decoding peaked at 170 ms, and position-invariant decoding

peaked at 180 ms after stimulus onset. Our findings add to this by investigating rotation-invariant decoding. We showed that fixed-rotation (non-invariant) object coding peaked ~116 ms after stimulus onset, whereas rotation-tolerant object coding peaked at ~192 ms after stimulus onset. Although it is not possible to directly compare the peak times across these different studies because of different stimuli and methods, our findings suggest that the rotation-tolerant representation is fully established later than the size-tolerant and position-tolerant representation. This would be in line with findings in non-human primates (Ratan Murty & Arun, 2017) as well as findings in humans at the category level (Karimi-Rouzbahani et al., 2017). Our findings therefore contribute to our understanding of how the human brain arrives at object representations that are tolerant to different transformations. In addition, our RSA method adds to previous findings by controlling for the contribution of low-level visual differences between stimuli. An object presented in different sizes, positions, or rotations will have consistent low-level visual features, such as luminance, which can be used by a classifier. The RSA method used in this study controls for the contribution of these features, allowing us to focus on high level representations that are tolerant to differences in rotation.

Previous work has provided insight into the time-course of the emergence of object information in the brain using fixed-rotation decoding methods, where the classifier is trained and tested on objects presented in the same (canonical) orientation (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019). While multivariate decoding methods are useful to study object representations in the human brain, one limitation is that differences in activation patterns caused by the different objects can be driven by low-level differences between objects. It is possible that there is information in perceptual areas about the low-level differences between objects, without having true object representations. The classifier cannot distinguish between these two possibilities and is likely to rely on both. By training and testing the classifier on different object rotations, we can force the classifier to generalise across rotations. This helps us tap into the higher-level

representations of objects that are tolerant to rotation, minimising the contribution of low-level differences in object features. Our results show a peak in rotation-tolerant object coding around 200 ms, which is approximately 80 ms later than the peak found for the fixed-rotation object decoding. The time-course of rotation-tolerant object information fits with that found for object decoding when training and testing on different exemplars of the same object in previous work (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019). A similar time-course has been found for category coding, which also shows a peak in decoding around 200 ms after stimulus onset (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Grootswagers, Robinson, & Carlson, 2019). This suggests that around this time a rich object representation has been established that is tolerant to rotation and generalises across different exemplars of the same object and category.

The time-course of decoding shows that the object representation persists after the presentation of the next stimulus in the sequence. If one assumes an afferent latency of around 70 ms to reach visual cortex, in line with when substantially above-chance decoding seems to emerge, then 200 ms later in the 5 Hz condition and 50 ms later in the 20 Hz condition, the next stimulus presented will have reached visual cortex, presumably interrupting any feedback loops (Mohsenzadeh et al., 2018). However, substantially above-chance decoding persists well beyond these times, which could be explained by a combination of the new stimulus not yet reaching particularly high level representations and possibly multiplexing, wherein successive stimuli are represented by distinct activation patterns that can coexist (King & Wyart, 2021; Marti & Dehaene, 2017). Our results add to the previous literature by showing that not only low-level features can persist, but that high-level invariant representations of different stimuli can be present in the brain simultaneously. Further, we showed that in the 20Hz condition, when processing was interrupted before the invariant representation was fully formed (at ~120 ms), multiple object representations were

maintained, but considerably less rotation-invariant information propagated through the system.

The decoding results suggest that the fixed-rotation object decoding method relies more on low-level visual differences between images compared to the rotation-tolerant approach. However, it is still possible that the classifier uses low-level differences for the rotation-tolerant image classification. We therefore used RSA to further minimise the effect of these low-level visual differences. Specifically, we correlated RDMs of the EEG data for the two different presentation rates with a model RDM for the rotation-tolerant object, while partialling out different low-level visual models. For the 5 Hz condition, the peak time of the partial correlation between the EEG and the rotation-tolerant model was ~184 ms after stimulus onset. This peak time is a similar time as the peak of rotation-tolerant decoding, suggesting higher-level object information contributed more strongly to the rotation-tolerant object decoding than low-level differences between images, as partialling out the low-level visual models did not impact the peak decoding time. However, when comparing the RSA and decoding methods for the comparison between the 5 Hz and 20 Hz conditions, our RSA findings revealed stronger partial correlations for the 5 Hz compared to 20 Hz condition starting only after ~164 ms. This is ~76 ms later than the difference between decoding accuracies for the 5 Hz and 20 Hz conditions found through the decoding analysis. This implies that there are no differences for the first peak (at ~120 ms) between the 5 Hz and 20 Hz conditions when the low-level visual information is partialled out. However, there is lower rotation-invariant information for the 20 Hz compared to 5 Hz condition in the second peak (at ~200ms). This suggests that faster presentation speeds limit higher-level object representations more than the representation of low-level visual information.

The cross-decoding indicates that the rotation-tolerant object representations for the fast and slow presentation rates are similar, although the asymmetry of some off-diagonal decoding suggests that the representation of rotation-tolerant object information was slightly

slower or later for the 5 Hz compared to the 20 Hz condition. This finding is in line with previous work (Grootswagers, Robinson, & Carlson, 2019; Robinson et al., 2019) that showed a disruption, but not elimination, of object and category processing when objects were presented at 20 Hz compared to 5 Hz. Together, these results suggest that although the faster presentation speed affects object processing, it does not fully disrupt the object processing.

One interesting aspect of the apparent transition from orientation-specific to rotation-tolerant representation is that there appears to be two distinct processing stages rather than a smooth transition. That is, one might have expected a gradual, smooth decrease in the fixed-rotation information as the rotation-tolerant information emerged. Instead, the data show two peaks in the fixed-rotation object decoding, as if two distinct processing stages are involved. This was also observed in some previous papers (Grootswagers, Robinson, & Carlson, 2019; Grootswagers, Robinson, Shatek, et al., 2019; Robinson et al., 2019). An interesting aspect of the temporal generalisation data is also consistent with the theory that there are two discrete processing stages. At about the same time as the decoding accuracy peaks, the temporal generalisation matrix shows two peaks of generalisation, which are separated by a short low generalisation interval at about 160 ms. By “low generalisation interval”, we mean that training on the EEG data at that time yields reasonable decoding accuracy but very little generalisation to other times, compared for instance to 220 ms, when the decoding accuracy is about the same but much more generalisation occurs. This is consistent with the brain’s representations at 160 ms being quite transient (not generalising to other times), a sort of transitional period as the rotation-tolerant information emerges. Perhaps the calculations required for the rotation-tolerant information require an intermediate representation or code that is almost immediately discarded.

Another interesting, although not reliable, result is the existence of negative correlations of the EEG data with the low-level image feature models, in particular for the colour, luminance, and silhouette models, but not for the mean luminance or RMS contrast

models. The correlations dropped off largely between 150 and 200 ms, which roughly corresponds to time between the first and second decoding peaks, when we have speculated that there is a short-lived transitional representation between the low-level and rotation-tolerant representations. Because the colour, luminance, and silhouette are all pixel-wise spatially specific models (and the mean luminance and RMS contrast models are not), it makes sense that their correlations would diminish and even become negative as the brain's representations move away from position-specific low-level representations. However, we do not know how to explain why the negative correlations are restricted to 150 to 200 ms, which roughly is the transitional period, before becoming positive again when the representations are most rotation-tolerant. Future studies could explore this in more detail.

Taken together, this study provides insight into the temporal dynamics of the emergence of rotation-tolerant object information in the brain. The time-course of rotation-tolerant object representations shows a peak around 200 ms for both the RSA and decoding analyses. This suggests that around this time the object information has been 'untangled' (DiCarlo & Cox, 2007), and a rich object representation has been established that is tolerant to rotation and generalises across different exemplars of the same object. These results highlight the importance of looking beyond the simple object decoding methods, which cannot separate the contribution of low-level feature processing from true invariant object representations.

5. ACKNOWLEDGEMENTS

This work was supported by an Australian Research Council (ARC) Discovery Early Career Researcher Award awarded to AKR (DE200101159), and ARC Discovery Projects awarded to TAC (DP160101300 and DP200101787). We would like to acknowledge the University of Sydney HPC service for providing High Performance Computing resources.

6. REFERENCES

Andresen, D. R., Vinberg, J., & Grill-Spector, K. (2009). The representation of object viewpoint in human visual cortex. *NeuroImage*, *45*(2), 522–536. https://doi.org/10.1016/j.neuroimage.2008.11.009

Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, *11*(10), 9–9. https://doi.org/10.1167/11.10.9

Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1. https://doi.org/10.1167/13.10.1

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Collins, E., Robinson, A. K., & Behrmann, M. (2018). Distinct neural processes for the perception of familiar versus unfamiliar faces along the visual hierarchy revealed by EEG. *NeuroImage*, *181*, 120–131. https://doi.org/10.1016/j.neuroimage.2018.06.080

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2017.02.013

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341. https://doi.org/10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How Does the Brain Solve Visual Object Recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, *32*(12), 2385–2400. https://doi.org/10.1016/0042-6989(92)90102-O

Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R. J., & Rees, G. (2008). fMRI Activity Patterns in Human LOC Carry Information about Object Exemplars within Category. *Journal of Cognitive Neuroscience*, *20*(2), 356–370. https://doi.org/10.1162/jocn.2008.20019

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex. *Neuron*, *24*(1), 187–203. https://doi.org/10.1016/S0896-6273(00)80832-6

Grootswagers, T., Robinson, A. K., & Carlson, T. A. (2019). The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage*, *188*, 668–679. https://doi.org/10.1016/j.neuroimage.2018.12.046

Grootswagers, T., Robinson, A. K., Shatek, S. M., & Carlson, T. A. (2019). Untangling featural and conceptual object representations. *NeuroImage*, *202*, 116083. https://doi.org/10.1016/j.neuroimage.2019.116083

Grootswagers, T., Robinson, A. K., Shatek, S. M., & Carlson, T. A. (2021). The neural dynamics underlying prioritisation of task-relevant information. *Neurons, Behavior, Data Analysis, and Theory*, 2020.06.25.172643. https://doi.org/10.51628/001c.21174

Guyonneau, R., Kirchner, H., & Thorpe, S. J. (2006). Animals roll around the clock: The rotation invariance of ultrarapid visual processing. *Journal of Vision*, *6*(10), 1–1. https://doi.org/10.1167/6.10.1

Harrison, W. J. (2022). Luminance and Contrast of Images in the THINGS Database. *Perception*, *51*(4), 244–262. https://doi.org/10.1177/03010066221083397

Holcombe, A. O. (2009). Seeing slow and seeing fast: two limits on perception. *Trends in Cognitive Sciences*, *13*(5), 216–221. https://doi.org/10.1016/j.tics.2009.02.005

Hubel, D. H., & Wiesel, T. N. (1963). Shape and arrangement of columns in cat's striate cortex. *The Journal of Physiology*, *165*(3), 559–568. https://doi.org/10.1113/jphysiol.1963.sp007079

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast Readout of Object Identity from Macaque Inferior Temporal Cortex. *Science*, *310*(5749), 863–866. https://doi.org/10.1126/science.1117593

Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*, *111*(1), 91–102. https://doi.org/10.1152/jn.00394.2013

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, *73*(1), 218–226. https://doi.org/10.1152/jn.1995.73.1.218

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., & Goodale, M. A. (2002). Differential Effects of Viewpoint on Object-Driven Activation in Dorsal and Ventral Streams. *Neuron*, *35*(4), 793–801. https://doi.org/10.1016/S0896-6273(02)00803-6

Jeffreys, H. (1998). *The Theory of Probability*. OUP Oxford.

Jolicoeur, P., & Milliken, B. (1989). Identification of Disoriented Objects: Effects of Context of Prior Presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 200–210. https://doi.org/10.1037/0278-7393.15.2.200

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, *115*(4), 2246–2250. https://doi.org/10.1152/jn.01074.2015

Karimi-Rouzbahani, H., Bagheri, N., & Ebrahimpour, R. (2017). Hard-wired feed-forward visual mechanisms of the brain compensate for affine variations in object recognition. *Neuroscience*, *349*, 48–63. https://doi.org/10.1016/j.neuroscience.2017.02.050

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

King, J.-R., & Wyart, V. (2021). The Human Brain Encodes a Chronicle of Visual Events at Each Instant of Time Through the Multiplexing of Traveling Waves. *Journal of Neuroscience*, *41*(34), 7224–7233. https://doi.org/10.1523/JNEUROSCI.2098-20.2021

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.004.2008

Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 395–416. https://doi.org/10.1037/0096-1523.22.2.395

Li, N., Cox, D. D., Zoccolan, D., & DiCarlo, J. J. (2009). What Response Properties Do Individual Neurons Need to Underlie Position and Clutter “Invariant” Object Recognition? *Journal of Neurophysiology*, *102*(1), 360–376. https://doi.org/10.1152/jn.90745.2008

Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nature Communications*, *8*(1), 1955. https://doi.org/10.1038/s41467-017-02079-x

McKeeff, T. J., Remus, D. A., & Tong, F. (2007). Temporal Limitations in Object Processing Across the Human Ventral Visual Pathway. *Journal of Neurophysiology*, *98*(1), 382–393. https://doi.org/10.1152/jn.00568.2006

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex. *Journal of Neurophysiology*, *100*(3), 1407–1419. https://doi.org/10.1152/jn.90248.2008

Mohsenzadeh, Y., Qin, S., Cichy, R. M., & Pantazis, D. (2018). Ultra-Rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *ELife*, *7*, e36329. https://doi.org/10.7554/eLife.36329

Morey, R. D., Romeijn, J.-W., & Rouder, J. N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, *72*, 6–18. https://doi.org/10.1016/j.jmp.2015.11.001

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs*. https://CRAN.R-project.org/package=BayesFactor

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*. https://doi.org/10.3389/fninf.2016.00027

Ratan Murty, N. A., & Arun, S. P. (2015). Dynamics of 3D view invariance in monkey inferotemporal cortex. *Journal of Neurophysiology*, *113*(7), 2180–2194. https://doi.org/10.1152/jn.00810.2014

Ratan Murty, N. A., & Arun, S. P. (2017). A Balanced Comparison of Object Invariances in Monkey IT Neurons. *Eneuro*, *4*(2), ENEURO.0333-16.2017. https://doi.org/10.1523/ENEURO.0333-16.2017

Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging Object Representations in the Visual System Predict Reaction Times for Categorization. *PLOS Computational Biology*, *11*(6), e1004316. https://doi.org/10.1371/journal.pcbi.1004316

Robinson, A. K., Grootswagers, T., & Carlson, T. A. (2019). The influence of image masking on object representations during rapid serial visual presentation. *NeuroImage*, *197*, 224–231. https://doi.org/10.1016/j.neuroimage.2019.04.050

Robinson, A. K., Grootswagers, T., Shatek, S. M., Gerboni, J., Holcombe, A., & Carlson, T. A. (2021). Overlapping neural representations for the position of visible and imagined objects. *Neurons, Behavior, Data Analysis, and Theory*, *4*(1), 1–28. https://doi.org/10.51628/001c.19129

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, *3*(6), 440–455. https://doi.org/10.1167/3.6.5

Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and Tolerance ("Invariance") Both Increase as Visual Information Propagates from Cortical Area V4 to IT. *Journal of Neuroscience*, *30*(39), 12978–12995. https://doi.org/10.1523/JNEUROSCI.0179-10.2010

Teichmann, L., Moerel, D., Baker, C. I., & Grootswagers, T. (2022). An empirically-driven guide on using Bayes Factors for M/EEG decoding. *Aperture Neuro*, *1*(8), 1–10.

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience*, *5*(5), 491–499. https://doi.org/10.1038/nn839

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on Psychological Science*, *6*(3), 291–298. https://doi.org/10.1177/1745691611406923

Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007). Trade-Off between Object Selectivity and Tolerance in Monkey Inferotemporal Cortex. *Journal of Neuroscience*, *27*(45), 12292–12307. https://doi.org/10.1523/JNEUROSCI.1897-07.2007