Untangling featural and conceptual object representations

Tijl Grootswagers[1,2,^], Amanda K. Robinson[1,2], Sophia M. Shatek[1], Thomas A. Carlson[1]

[1] School of Psychology, University of Sydney, Sydney, NSW, Australia

[2] Perception in Action Research Centre, Macquarie University, Sydney, NSW, Australia

[^] Corresponding author: tijl.grootswagers@sydney.edu.au

Abstract

How are visual inputs transformed into conceptual representations by the human visual system? The contents of human perception, such as objects presented on a visual display, can reliably be decoded from voxel activation patterns in fMRI, and in evoked sensor activations in MEG and EEG. A prevailing question is the extent to which brain activation associated with object categories is due to statistical regularities of visual features within object categories. Here, we assessed the contribution of mid-level features to conceptual category decoding using EEG and a novel fast periodic decoding paradigm. Our study used a stimulus set consisting of intact objects from the animate (e.g., fish) and inanimate categories (e.g., chair) and scrambled versions of the same objects that were unrecognizable and preserved their visual features (Long, Yu, & Konkle, 2018). By presenting the images at different periodic rates, we biased processing to different levels of the visual hierarchy. We found that scrambled objects and their intact counterparts elicited similar patterns of activation, which could be used to decode the conceptual category (animate or inanimate), even for the unrecognizable scrambled objects. Animacy decoding for the scrambled objects, however, was only possible at the slowest periodic presentation rate. Animacy decoding for intact objects was faster, more robust, and could be achieved at faster presentation rates. Our results confirm that the mid-level visual features preserved in the scrambled objects contribute to animacy decoding, but also demonstrate that the dynamics vary markedly for intact versus scrambled objects. Our findings suggest a complex interplay between visual feature coding and categorical representations that is mediated by the visual system's capacity to use image features to resolve a recognisable object.

Introduction

How does the brain transform perceptual information into meaningful concepts and categories? One key organisational principle of object representations in the human ventral temporal cortex is animacy (Caramazza & Mahon, 2003; Caramazza & Shelton, 1998; Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte et al., 2008; Mahon & Caramazza, 2011; Spelke, Phillips, & Woodward, 1995). Operationalised as objects that can move on their own volition, animate objects evoke different activation patterns than inanimate objects in human brain activity patterns in fMRI (Cichy, Pantazis, & Oliva, 2014; Connolly et al., 2012; Downing, Jiang, Shuman, & Kanwisher, 2001; Grootswagers, Cichy, & Carlson, 2018; Konkle & Caramazza, 2013; Kriegeskorte et al., 2008) and in MEG/EEG (Carlson, Tovar, Alink, & Kriegeskorte, 2013; Contini, Wardle, & Carlson, 2017; Grootswagers, Ritchie, Wardle, Heathcote, & Carlson, 2017; Grootswagers, Robinson, & Carlson, 2019; Kaneshiro, Guimaraes, Kim, Norcia, & Suppes, 2015; Ritchie, Tovar, & Carlson, 2015). A current theoretical debate concerns the degree to which categorical object representations in ventral temporal cortex are due to systematic featural differences within categories (Long et al., 2018; op de Beeck, Haushofer, & Kanwisher, 2008; Proklova, Kaiser, & Peelen, 2016).

Recent work has focused on understanding the contribution of visual features to the brain's representation of categories, such as animacy. This work has shown that a substantial proportion of animacy (de)coding in ventral temporal cortex can be explained by low and mid-level visual features (e.g., texture and curvature) that are inherently associated with animate versus inanimate objects (Andrews, Watson, Rice, & Hartley, 2015; Bracci & Op de Beeck, 2016; Bracci, Ritchie, & de Beeck, 2017; Bracci, Ritchie, Kalfas, & Op de Beeck, 2019; Coggan, Liu, Baker, & Andrews, 2016; Kaiser, Azzalini, & Peelen, 2016; Long et al., 2018; Proklova et al., 2016; Rice, Watson, Hartley, & Andrews, 2014; Ritchie, Bracci, & op de Beeck, in press; Watson, Young, & Andrews, 2016). Long et al. (2018) recently investigated how mid-level features contribute to categorical representations using images of intact objects and scrambled "texform" versions of the same objects. Crucially, the texform versions of the objects were unrecognisable (at the individual image identity level) but preserved mid-level features such as texture. Using fMRI, they found the categories of animacy and size were similarly coded in the brain for intact and texform versions of objects, thus demonstrating that such patterns can arise without the explicit recognition of an object (Long et al., 2018). In MEG and EEG, one study showed that animate and inanimate objects cannot be

differentiated when they are closely matched for shape (Proklova, Kaiser, & Peelen, 2019). Other studies, however, have found that object animacy decoding generalises to unseen exemplars with different shapes (cf. Contini et al., 2017), suggesting animacy decoding, in part, might be based on general conceptual representations. Taken together, these results suggest that either there is some abstract conceptual representation of animacy, or that objects within the animate and inanimate categories share sufficient visual regularities to drive the categorical organisation of object representations in the brain.

In the current study, we tested the contribution of visual features to the dynamics of emerging conceptual representations. We used a previously published stimulus set (Figure 1) that was designed to test the contribution of mid-level features to conceptual categories (animacy and size) in the visual system (Long et al., 2018), which consisted of luminance-matched real objects, and scrambled, "texform" versions of the same objects that retain mid-level texture and form information (Long, Störmer, & Alvarez, 2017; Long et al., 2018). We used EEG and a rapid-MVPA paradigm (Grootswagers et al., 2019) to study the emergence of conceptual information. Based on previous fMRI work (Long et al., 2018), we predicted that texforms would evoke animacy-like patterns in the EEG signal similar to intact objects. In addition, we hypothesized that animacy-like patterns evoked by texforms may need more time to develop. To test this, we presented the stimuli at varying rapid presentation rates, as faster rates have been shown to limit the depth of stimulus processing (Collins, Robinson, & Behrmann, 2018; Grootswagers et al., 2019; McKeeff, Remus, & Tong, 2007; Robinson, Grootswagers, & Carlson, 2019). We found that EEG activation patterns of texform versions of the objects were decodable, but that conceptual categorical decoding of intact objects was more robust, and could be achieved at faster presentation rates, which suggests that the visual system needs less time to process the intact objects. Together, our results provide evidence that visual features contribute to the representation of conceptual object categories, but also show that higher level abstractions cannot be fully explained by statistical regularities.

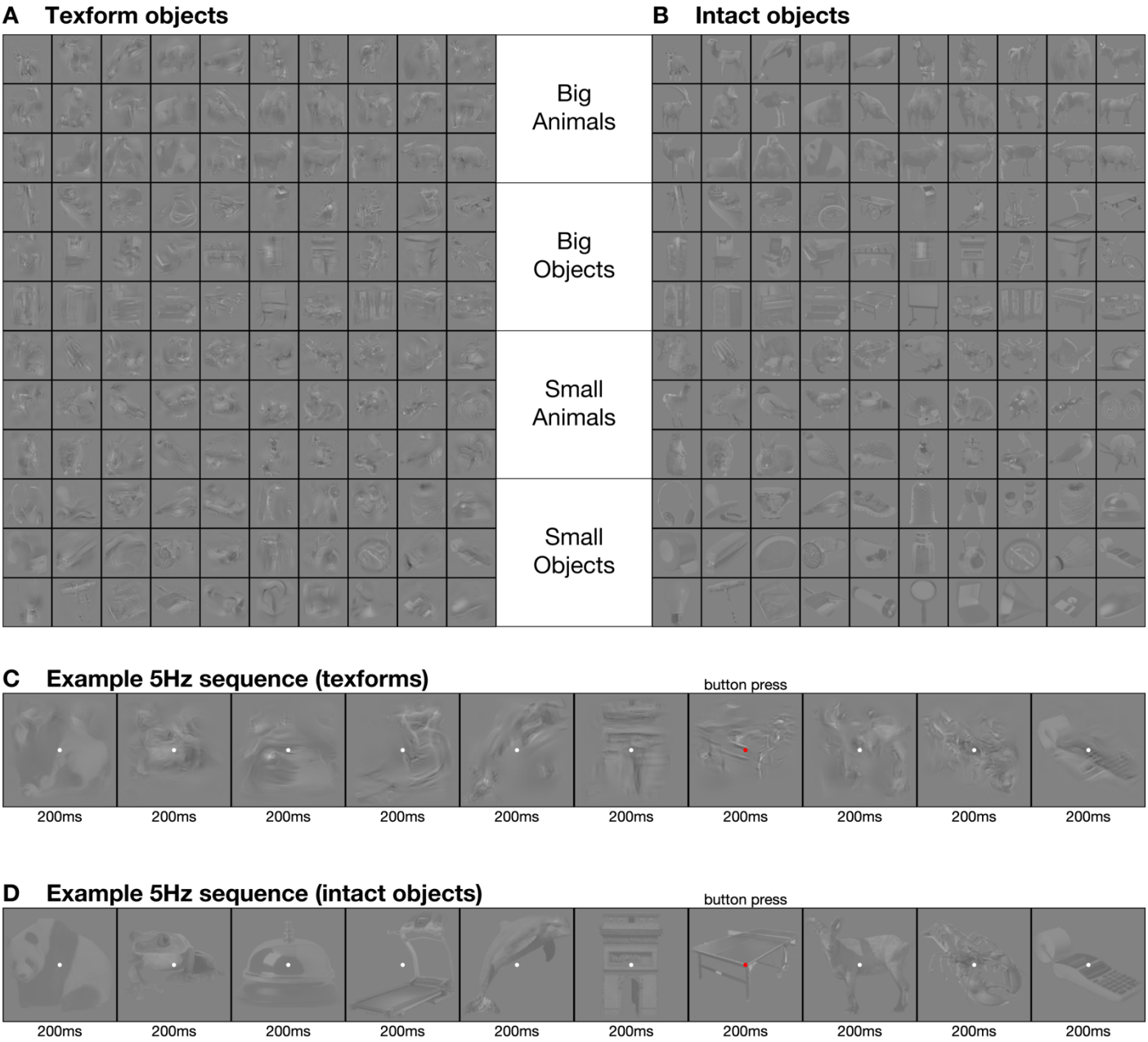

Figure 1. Stimuli and design. Stimuli were 120 objects categorizable as animate or inanimate, and as big or small. A. The first half of the experiment used texform versions of objects (presented first so that participants were not aware of their intact counterparts). B. In the second half of the experiment, the original intact versions were used. All images were obtained from https://osf.io/69pbd/ (Long et al., 2018). Stimuli were presented at four presentation frequencies. C. Example texform sequence at 5Hz, where stimuli were presented for 200ms each. D. Example intact object sequence at 5Hz. The sequence presentation orders for intact objects and texforms were matched. Participants performed an orthogonal task where they responded with a button press to the fixation dot turning red.

Methods

Stimuli, data, and analysis code are available online through https://osf.io/sz9ve.

Participants

Participants were 20 volunteers (11 females, 9 males; mean age 24.6, age range: 17-59) recruited from the University of Sydney in return for payment or course credit. All participants reported normal or corrected-to-normal vision. Two participants were left-handed. The study was approved by the University of Sydney ethics committee and informed consent in writing was obtained from all participants.

Stimuli and design

Stimuli were obtained from https://osf.io/69pbd (Long et al., 2018). For a full description of the stimulus generation procedures, see Long et al., 2018. The stimuli were 120 visual objects that were grouped in four categories: big animals, small animals, big objects, and small objects. This allowed for orthogonal animacy and size categorisation of the stimuli. All stimuli were matched for average luminance. The stimuli underwent a scrambling procedure (Long et al., 2018) to generate texform versions of the same objects. All 240 stimuli, 120 intact objects, and 120 texform versions were used in this experiment (Figure 1).

Following the procedure of Long et al., we presented participants with texform versions of the stimuli in the first half of the experiment, and with intact objects in the second half of the experiment. Participants were all naïve to the experiment aims and were not informed about the relationship between the texforms and intact images. We used a rapid serial visual processing paradigm to present the stimuli in fast succession (Grootswagers et al., 2019). Stimuli were presented in random order in streams at four presentation frequencies: 60Hz, 30Hz, 20Hz, and 5Hz, always using a 100% duty cycle, following previous work that investigated category decoding at fast presentation rates (Grootswagers et al., 2019; Mohsenzadeh, Qin, Cichy, & Pantazis, 2018). All stimuli within a category (texforms/objects) were presented in each stream (i.e., every stream contained 120 images). Stimuli were presented at 6.8 x 6.8 degrees of visual angle on a grey background and were overlaid with a white fixation dot of 0.2 degrees diameter (Figure X). During the experiment, participants responded with a button press when the dot changed

colour, which happened between 1 and 4 times during each stream, at random positions in the stream. Each object was presented 30 times in each condition (intact and texform), and at each presentation frequency. The experiment lasted about 40 minutes.

EEG recordings and preprocessing

Continuous EEG data were recorded from 64 electrodes arranged according to the international standard 10–10 electrode placement system (Oostenveld & Praamstra, 2001) using a BrainVision ActiChamp system, digitized at a 1000-Hz sample rate and referenced online to Cz. Preprocessing was performed offline using EEGlab (Delorme & Makeig, 2004). Data were filtered using a Hamming windowed FIR filter with 0.1Hz highpass and 100Hz lowpass filters and were downsampled to 250Hz. No further preprocessing steps were applied. All analyses were performed on the channel voltages at each time point. Epochs were created for each stimulus presentation ranging from [-100 to 1000 ms] relative to stimulus onset.

Decoding analysis

We applied an MVPA decoding pipeline (Grootswagers, Wardle, & Carlson, 2017) applied to the EEG channel voltages. The decoding analyses were implemented in CoSMoMVPA (Oosterhof, Connolly, & Haxby, 2016). Regularised linear discriminant analysis (LDA) classifiers were used in combination with a sequence cross-validation approach to decode pairwise image identities firstly between all pairs of texform images, and secondly between all pairs of intact object images. For animacy decoding, an exemplar-by-sequence cross-validation approach was used (Carlson et al., 2013; Grootswagers et al., 2019). That is, a pair of animate and inanimate images from one sequence was used as test data, and classifiers were trained on the remaining images from the remaining sequences. This was repeated for all animate-inanimate pairs and all sequences, averaging the resulting cross-validated prediction accuracies. Real-world size decoding used the same pipeline, with an exemplar-by-sequence cross-validation procedure. To test for the similarity between texform and object patterns, we performed the same analyses and cross-validation designs described above, but we trained the classifiers on intact object sequences and tested on the texform sequences, and vice versa. All analyses were repeated for each time point in the epochs,

resulting in a decoding accuracy over time for every presentation frequency, within subject. The subject-averaged results for each frequency were analysed at the group level.

Statistical inference

For each decoding analysis, we used Bayesian statistics to determine the evidence for above chance decoding or non-zero differences between texform and intact object decoding accuracies (Dienes, 2011, 2016; Jeffreys, 1961; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007). For the alternative hypothesis of above-chance (50%) decoding or a non-zero difference, a JZS prior (Rouder et al., 2009) was set with a scale factor of 0.707 (Jeffreys, 1961; Rouder et al., 2009; Wetzels & Wagenmakers, 2012; Zellner & Siow, 1980). We then calculated the Bayes factor (BF) which is the probability of the data under the alternative hypothesis relative to the null hypothesis. We thresholded BF > 10 as evidence for the alternative hypothesis, and BF < 1/3 as evidence in favour of the null hypothesis (Jeffreys, 1961; Wetzels et al., 2011). In addition to the Bayes factors, we computed p-values for decoding against chance, and the differences between texform and intact object decoding accuracies. We used a sign-swap permutation test (1000 iterations), and computed threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009) values at each time point. To correct for multiple comparisons, the maximum TFCE statistic across time for each permutation were selected to form a corrected null-distribution (Maris & Oostenveld, 2007). We then calculated p-values by comparing the observed TFCE values to the corrected permutation distribution.

Exploratory channel-searchlight analysis

To obtain insights into the source of the difference between texform and intact object decoding, we performed a channel by timepoint searchlight. For all contrasts, we performed multiclass decoding instead of all pairwise comparisons to reduce computation time. A leave-one-sequence-out cross-validation approach was performed on local clusters of channels. For each channel, a local cluster was constructed by taking the closest four neighbouring channels, and the decoding analyses were performed on the signal of just these channels. The decoding accuracies were stored at the centre channel of the cluster. This resulted in a time by channel map of decoding accuracy for

each of the contrasts, and for each subject, at each frequency. Here, we reported the results for the decoding differences at 5Hz and have added the other frequencies to the project's online repository.

Results

Participants (N=20) viewed streams of texform stimuli and intact objects (Figure 1). The stimuli were presented in random order at four presentation frequencies (60Hz, 30Hz, 20Hz, 5Hz) to target different levels of visual processing (Grootswagers et al., 2019; Robinson et al., 2019). The stimuli were developed by Long et al., (2017), and obtained from https://osf.io/69pbd/ (Long et al., 2017, 2018). Continuous EEG was recorded during the streams and cut into overlapping epochs based on the onset of each stimulus within the streams. The epoched data were subjected to a multivariate decoding analysis, similar to previous work that decoded individual images in fast presentation streams (Grootswagers et al., 2019; Robinson et al., 2019).

To investigate how image representations differed between the texform and intact versions, we obtained cross-validated classifier performance between all pairwise texform images (Figure 2A), and all pairwise intact object images (Figure 2B). The differences between texform and intact object decoding accuracies (Figure 2C) showed evidence for no difference in the initial response (up to around 150ms), but higher accuracies for intact objects after that. These differences were localised in occipito-temporal areas (Figure 2D). Both texforms and intact objects were decodable from around 90 ms after stimulus onset at all presentation frequencies, characteristic of early stages of visual processing (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017). Faster presentation frequencies resulted in lower peak decoding and shorter decoding durations, consistent with previous results showing that fast rates restrict visual processing (Robinson et al., 2019). In general, the image-level decoding results were similar between texforms and intact objects, apart from the intact objects at 5Hz, where a larger second peak was observed that was not apparent for the texforms. To further investigate the similarity between the underlying patterns, we performed cross-decoding where we trained on intact object patterns and tested on texform versions (Figure 2E), and vice versa (Figure 2F). The results of these analyses showed that the evoked patterns are sufficiently similar to allow for above-chance cross-decoding at all presentation frequencies.

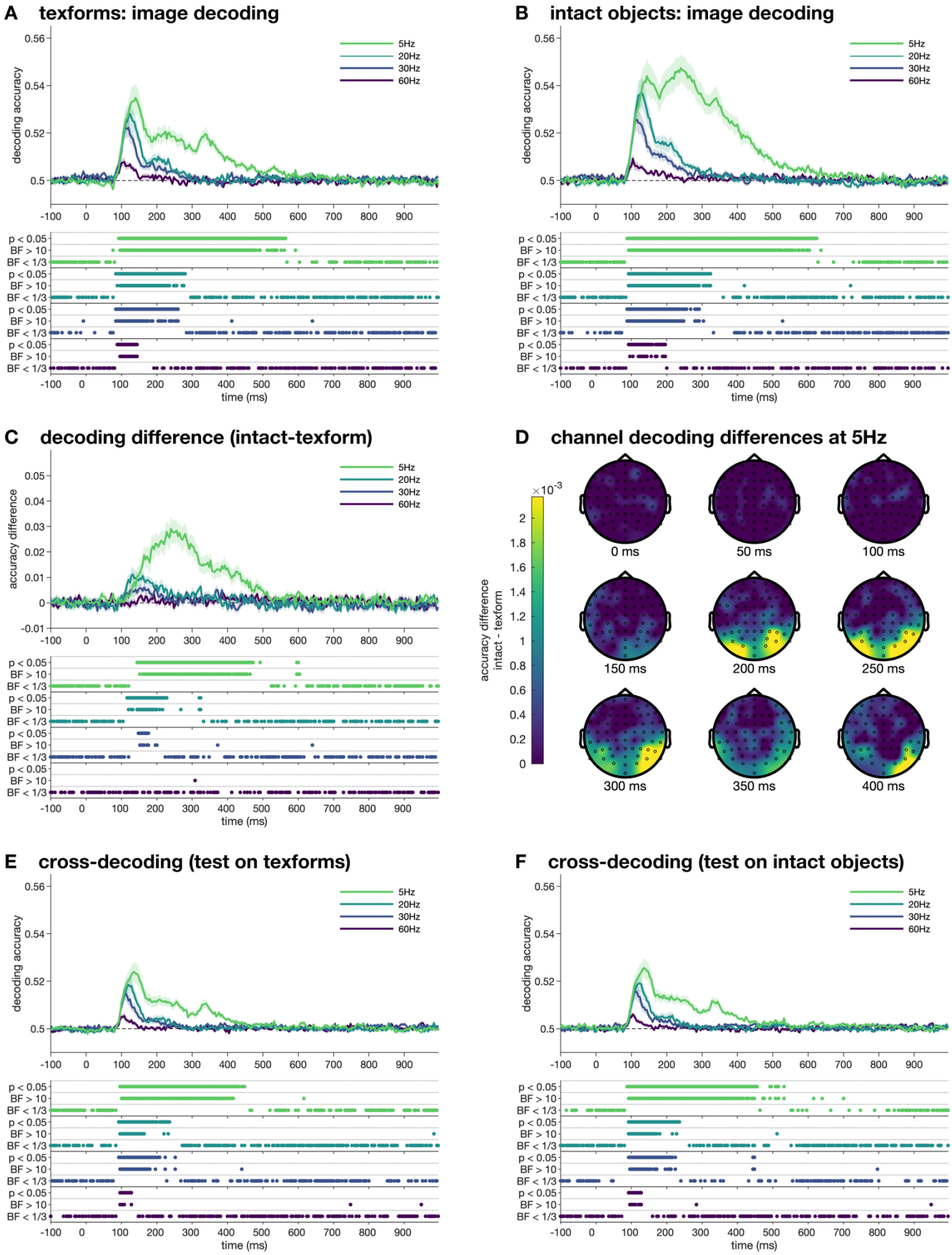

Figure 2. Decoding texforms and intact objects. A. Decoding between all texform image pairs. **B.** Decoding between all intact object image pairs. **C.** Difference between intact objects and texforms at each frequency. **D.** Channel searchlight maps at nine time points for the decoding differences at 5Hz. **E.** Decoding texforms, training the classifier on intact objects. **F.** Decoding intact objects, training on texforms. Different lines in each plot show decoding accuracy for different presentation frequencies over time relative to stimulus onset, with shaded areas showing standard error across subjects (N=20). Thresholded Bayes factors (BF) and p-values for above-chance decoding or non-zero differences are displayed below the plot for each frequency.

To investigate to what extent the visual features preserved in the texform versions of the objects drive categorical distinctions in activation patterns, we trained classifiers on decoding animacy from the stimuli, using an exemplar-by-sequence cross-validation approach to avoid overfitting to individual images (Carlson et al., 2013; Grootswagers et al., 2019). For texforms, above chance decoding of (featural) animacy was observed between 300ms and 400ms after stimulus onset (Figure 3A) for the 5Hz presentation frequency but was not evident for the faster presentation rates. Animacy decoding in intact object images, in contrast, was above chance for 5Hz, 20Hz and 30Hz (Figure 3B). Onset of animacy decoding was approximately 150ms for the 5Hz and 20Hz conditions, and at 220ms for the 30Hz frequency. The differences between texform and intact object animacy decoding accuracies (Figure 3C) highlight the substantial difference at 5Hz, which an exploratory searchlight suggested to be mainly located across occipito-temporal areas (Figure 3D). Cross-decoding (Figure 3E&F) showed that part of the animacy pattern generalised between intact objects and texforms. This result shows that the shared visual features between texforms and intact objects contributes to, but do not wholly explain, the categorical representation of animacy in the brain.

In the final analysis, we asked if the categorical representation of real-world size emerges similarly for intact and texform versions of objects. An exemplar-by-sequence cross-validation approach was used to decode real world size (small versus large objects) for the texform and intact objects. At none of the presentation rates was (featural) real-world size decodable from the texform stimuli (Figure 4A). Real world size of the intact object images was decodable for 5Hz and 20Hz frequencies (Figure 4B). The differences between texform and intact object size decoding accuracies are shown in Figure 4C. Cross-decoding showed evidence for no shared pattern of object size between intact objects and texforms (Figure 4E&F). Combined, the animacy and size decoding results show a fundamental difference in how conceptual categories emerge for intact objects and their scrambled counterparts.

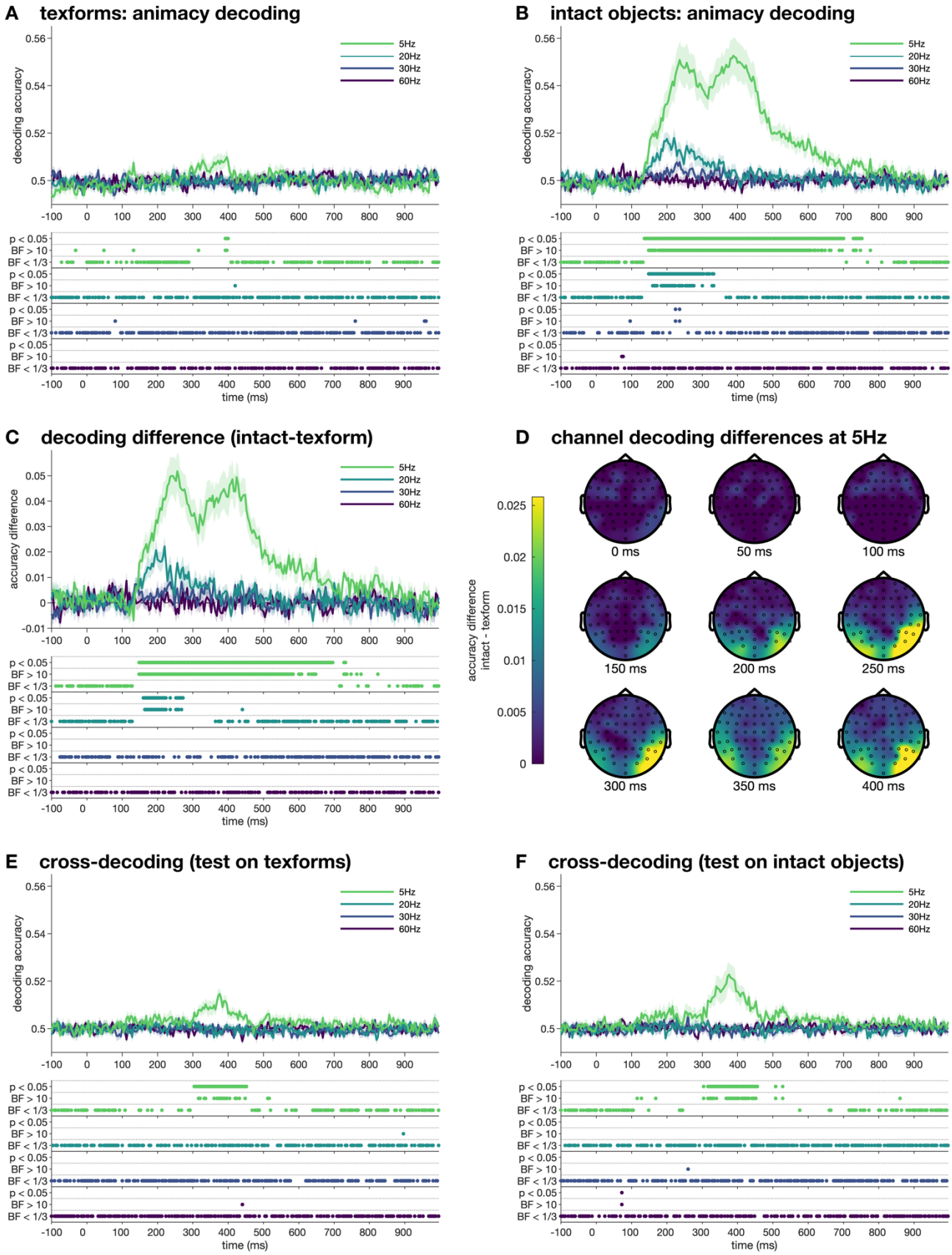

Figure 3. Decoding animacy from texforms and intact objects. A. Decoding the animacy of the texform images. **B.** Decoding the animacy of the intact objects. **C.** Difference between intact objects and texforms at each frequency. **D**. Channel searchlight maps at nine time points for the decoding differences at 5Hz. **E**. Decoding animacy from texforms, training the classifier on intact objects. **F.** Decoding animacy from intact objects, training on texforms. Different lines in each plot show decoding accuracy for different presentation frequencies over time relative to stimulus onset, with shaded areas showing standard error across subjects (N=20). Thresholded Bayes factors (BF) and p-values for above-chance decoding or non-zero differences are displayed below the plot for each frequency.

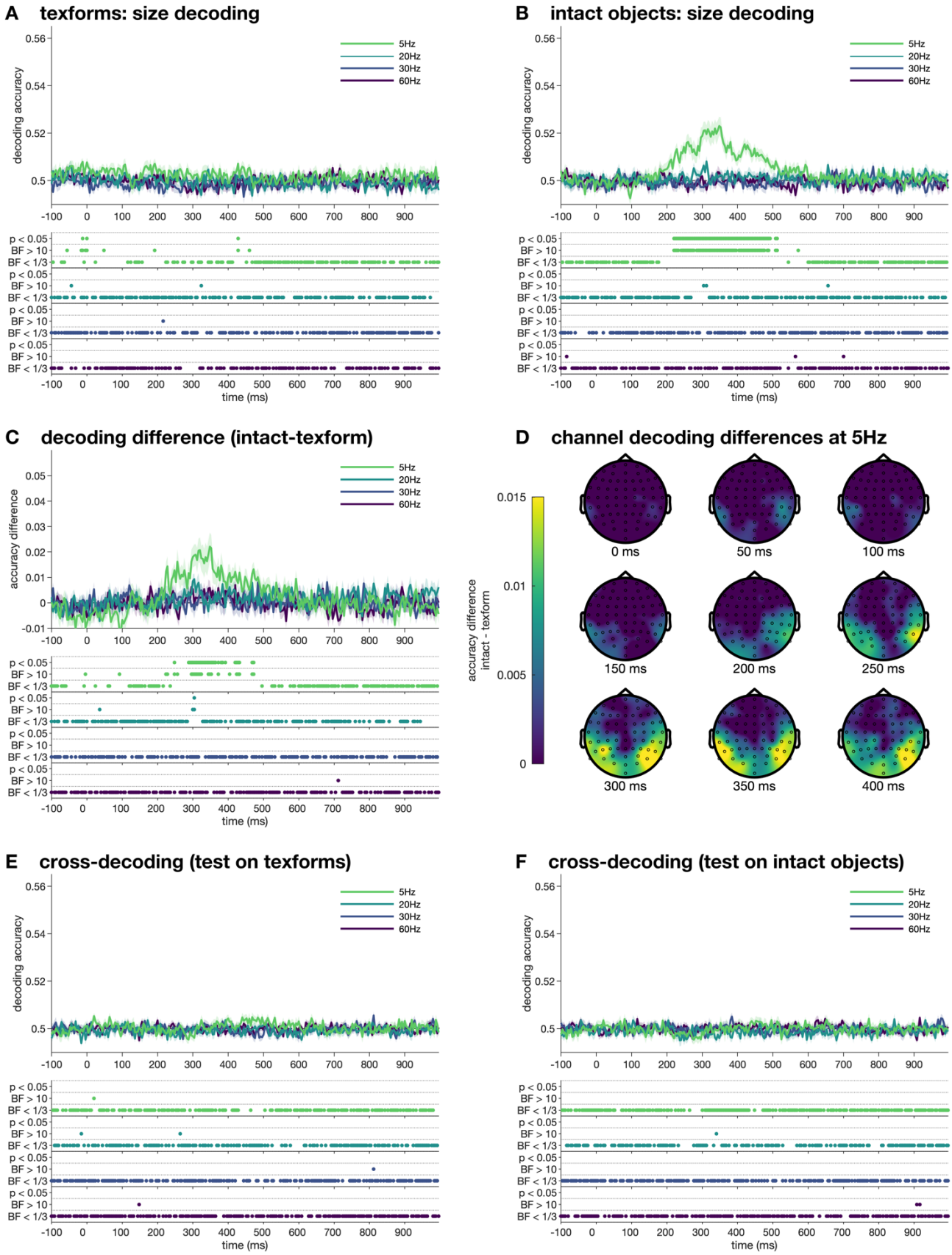

Figure 4. Size was only decodable from intact objects. A. Decoding the real-world size of the texform images. **B.** Decoding the real-world size of the intact objects. **C.** Difference between intact objects and texforms at each frequency. **D**. Channel searchlight maps at nine time points for the decoding differences at 5Hz. **E**. Decoding size from texforms, training the classifier on intact objects. **F.** Decoding size from intact objects, training on texforms. Different lines in each plot show decoding accuracy for different presentation frequencies over time relative to stimulus onset, with shaded areas showing standard error across subjects (N=20). Thresholded Bayes factors (BF) and p-values for above-chance decoding or non-zero differences are displayed below the plot for each frequency.

Discussion

In this study, we assessed the contribution of mid-level features to high level categorical object representations using a combination of fast periodic visual processing streams and multivariate EEG decoding. We used images of intact and texform versions of objects from a previously published study (Long et al., 2018) and found that their neural representations were similarly distinct at the image level. In contrast, the decoding accuracies of the original categorical distinctions of animacy and real-world size varied markedly across the texform and intact versions of the objects. The patterns of neural activity evoked by animate and inanimate intact objects were decodable during a larger time period than their texform versions, suggesting the temporal dynamics of animacy-like patterns varied between intact and texform versions despite their shared mid-level visual features. In addition, the animacy of intact objects was decodable at 5Hz, 20Hz and 30Hz, but texforms were only decodable at 5Hz. Higher level categorical brain regions exhibit larger responses to slower presentation rates relative to faster rates (McKeeff et al., 2007), and we previously found that slower object presentations reached higher, more abstract levels of visual processing (Grootswagers et al., 2019; Robinson et al., 2019). Thus, the absence of animacy decoding for texform objects at faster presentation rates indicates that higher level processing was required for the animate/inanimate distinction in texform stimuli. Moreover, a clear double-peak structure was observed for decoding the intact objects at 5Hz, but not for the texforms. This could reflect an additional conceptual processing step that is unique to intact objects presented at a frequency that allows reaching conceptual processing stages. We interpret these findings as evidence that shared visual features between texforms and intact objects contribute to, but do not wholly explain, the categorical organisation of animacy in the brain.

Our results corroborate fMRI results using the same stimuli showing that texforms and intact objects generated similar categorical representations along the visual hierarchy but that the recognizable images generated stronger category responses (Long et al., 2018). The current results further show a clear difference in the temporal dynamics of animacy representations within the visual system for featural versus conceptual object representations. At faster presentation rates, animacy decoding was observed in intact objects but not in texforms, indicating that intact objects promote categorical representations with limited processing. It is important to note that the intact objects were shown only in the second half of the experiment, which might have contributed to better animacy decoding for intact objects. However, image-level results were similar between texforms and intact objects, which suggests

that the experimental paradigm was not wholly responsible for the differences in categorical-level decoding. Together, these results suggest that brain responses to intact objects contain additional animacy category information over and above the statistical visual regularities present in the texforms.

The texform scrambling process was used to render images unrecognisable at the individual image level, while maintaining featural image statistics. Some low-level visual information may have been lost in the scrambling process, such as shape and curvature information, which is a strong cue for animacy (Levin, Takarae, Miner, & Keil, 2001; Schmidt, Hegele, & Fleming, 2017; Zachariou, Giacco, Ungerleider, & Yue, 2018). In MEG and EEG decoding studies, classification can be strongly driven by differences in object shape (Proklova et al., 2019), and silhouette similarity is often a strong predictor of the similarities between the earliest neural responses (Carlson et al., 2013; Grootswagers et al., 2019; Teichmann, Grootswagers, Carlson, & Rich, 2018; Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016). It is also important to note that while the texform images are not recognisable at the individual level, they can still be categorised (e.g., for animacy) above chance (Long et al., 2017). Human categorisation accuracies on these images was found to be predicted by the amount of curvilinear and rectilinear information in the image (Zachariou et al., 2018). Yet, even if intermediate visual features are sufficient to classify conceptual categories above-chance behaviourally, our results suggest that this is only possible given sufficient processing time.

These findings support the notion that large-scale categorical organisations in the visual system are to some extent driven by mid-level visual features. However, if concepts were decodable using only brain responses to mid-level feature, then this would predict above-chance decoding of concepts also at faster frequencies for the texforms. This was not the case in our results. Instead, we only observed animacy decoding for the slowest (5Hz) presentation frequency, which suggests that the conceptual animacy category only emerges from mid-level features after deeper processing. Thus, it could be the case that mid-level feature coding in early visual areas does not allow for concept decoding, but these features are "untangled" by higher visual areas into linearly separable categorical organisations (DiCarlo & Cox, 2007). This would mean that visual features could indeed drive the organisation in high level areas, but only given sufficient processing time for such untangling processes to complete. Furthermore, the speed of processing or information transfer to these higher visual areas could be modulated by the amount of evidence that supports the successful recognition of an object. For example, the intact objects have a well-defined

outline that separates the object from the background, while the edges of the texforms are more blurred. This potentially could disrupt segmentation processes which in turn delays the amount of time it takes for information to reach higher level recognition stages. The results of this study therefore suggest a complex interplay between early and late stages of processing that ultimately manifests in more abstract categorical representations.

A limitation of the current study is that the animacy and size category boundaries of the stimulus set also define four subcategories, for example, most small objects are tools, and most big animals are mammals. Therefore, the results could be driven by these subcategories, rather than overall conceptual animacy and size organisations. Future work could explore this possibility, using stimulus sets that in addition match the subcategories within animacy and size categories. In addition, while it is important to disentangle perceptual features from conceptual representations, the two are inherently intermingled. Categorical organisations, such as animacy, are strongly represented partly because they share perceptual characteristics, which makes them easier to discriminate. Indeed, inanimate stimuli that share perceptual features with animate items evoke brain responses that are similar to other animate stimuli (Bracci et al., 2019). On the other hand, neural responses to animate stimuli that share characteristics with inanimate objects (e.g., a starfish) are more confusable with inanimate stimuli (Grootswagers, Ritchie, et al., 2017). Moreover, when stimuli are closely matched in shape, the activation patterns can become indistinguishable (Proklova et al., 2016, 2019). Together, these examples could be taken to suggest that the dichotomy of animacy should be revised to more closely reflect, for example, a continuous account of perceptual and conceptual animal typicality (Connolly et al., 2012; Contini, Goddard, Grootswagers, Williams, & Carlson, 2019; Grootswagers, Ritchie, et al., 2017; Iordan, Greene, Beck, & Fei-Fei, 2016; Sha et al., 2015; Thorat, Proklova, & Peelen, 2019).

In conclusion, we found that animacy was decodable from texform versions of objects, but that animacy of intact objects was more strongly decodable, and at faster presentation frequencies. Information contained in the texform versions of the objects thus not fully account for the distinct patterns of neural responses evoked by conceptual object categories. These findings suggest that complex interactions between lower and higher levels of visual processing mediate the representations of category, which has important implications for disentangling perceptual and conceptual representations in the human brain.

Acknowledgements

This research was supported by an Australian Research Council Future Fellowship (FT120100816) and an Australian Research Council Discovery project (DP160101300) awarded to T.A.C. The authors acknowledge the University of Sydney HPC service for providing High Performance Computing resources. The authors declare no competing financial interests.

References

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision*, *15*(7), 3–3. https://doi.org/10.1167/15.7.3

Bracci, S., & Op de Beeck, H. P. (2016). Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. *Journal of Neuroscience*, *36*(2), 432–444. https://doi.org/10.1523/JNEUROSCI.2314-15.2016

Bracci, S., Ritchie, J. B., & de Beeck, H. O. (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia*, *105*, 153–164. https://doi.org/10.1016/j.neuropsychologia.2017.06.010

Bracci, S., Ritchie, J. B., Kalfas, I., & Op de Beeck, H. (2019). The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *Journal of Neuroscience*, 1714–1718. https://doi.org/10.1523/JNEUROSCI.1714-18.2019

Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, *7*(8), 354–361. https://doi.org/10.1016/S1364-6613(03)00159-1

Caramazza, A., & Shelton, J. R. (1998). Domain-Specific Knowledge Systems in the Brain: The Animate-Inanimate Distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. https://doi.org/10.1162/089892998563752

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

Coggan, D. D., Liu, W., Baker, D. H., & Andrews, T. J. (2016). Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information. *NeuroImage*, *135*, 107–114. https://doi.org/10.1016/j.neuroimage.2016.04.060

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

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., … Haxby, J. V. (2012). The Representation of Biological Classes in the Human Brain. *The Journal of Neuroscience*, *32*(8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

Contini, E. W., Goddard, E., Grootswagers, T., Williams, M., & Carlson, T. (2019). A humanness dimension to visual object coding in the brain. *BioRxiv*, 648998. https://doi.org/10.1101/648998

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

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

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Grootswagers, T., Cichy, R. M., & Carlson, T. A. (2018). Finding decodable information that can be read out in behaviour. *NeuroImage*, *179*, 252–262. https://doi.org/10.1016/j.neuroimage.2018.06.022

Grootswagers, T., Ritchie, J. B., Wardle, S. G., Heathcote, A., & Carlson, T. A. (2017). Asymmetric Compression of Representational Space for Object Animacy Categorization under Degraded Viewing Conditions. *Journal of Cognitive Neuroscience*, *29*(12), 1995–2010. https://doi.org/10.1162/jocn_a_01177

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., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2016). Typicality sharpens category representations in object-selective cortex. *NeuroImage*, *134*, 170–179. https://doi.org/10.1016/j.neuroimage.2016.04.012

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.

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

Kaneshiro, B., Guimaraes, M. P., Kim, H.-S., Norcia, A. M., & Suppes, P. (2015). A Representational Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification. *PLOS ONE*, *10*(8), e0135697. https://doi.org/10.1371/journal.pone.0135697

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309. https://doi.org/10.1152/jn.00024.2007

Konkle, T., & Caramazza, A. (2013). Tripartite Organization of the Ventral Stream by Animacy and Object Size. *Journal of Neuroscience*, *33*(25), 10235–10242. https://doi.org/10.1523/JNEUROSCI.0983-13.2013

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics*, *63*(4), 676–697.

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, *17*(6), 20–20. https://doi.org/10.1167/17.6.20

Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences*, 201719616. https://doi.org/10.1073/pnas.1719616115

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences*, *15*(3), 97–103. https://doi.org/10.1016/j.tics.2011.01.004

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

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

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

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

op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience*, *9*(2), 123–135. https://doi.org/10.1038/nrn2314

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. *Journal of Cognitive Neuroscience*, 1–13. https://doi.org/10.1162/jocn_a_00924

Proklova, D., Kaiser, D., & Peelen, M. V. (2019). MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2019.03.028

Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-Level Image Properties of Visual Objects Predict Patterns of Neural Response across Category-Selective Regions of the Ventral Visual Pathway. *The Journal of Neuroscience*, *34*(26), 8837–8844. https://doi.org/10.1523/JNEUROSCI.5265-13.2014

Ritchie, J. B., Bracci, S., & op de Beeck, H. P. (in press). Avoiding illusory effects in representational similarity analysis: What (not) to do with the diagonal. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2016.12.079

Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging Object Representations in the Visual System Predict Reaction Times for Categorization. *PLoS Comput Biol*, *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

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.

Schmidt, F., Hegele, M., & Fleming, R. W. (2017). Perceiving animacy from shape. *Journal of Vision*, *17*(11), 10–10. https://doi.org/10.1167/17.11.10

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The Animacy Continuum in the Human Ventral Vision Pathway. *Journal of Cognitive Neuroscience*, *27*(4), 665–678. https://doi.org/10.1162/jocn_a_00733

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants' knowledge of object motion and human action. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). New York, NY, US: Clarendon Press/Oxford University Press.

Teichmann, L., Grootswagers, T., Carlson, T., & Rich, A. N. (2018). Decoding Digits and Dice with Magnetoencephalography: Evidence for a Shared Representation of Magnitude. *Journal of Cognitive Neuroscience*, *30*(7), 999–1010. https://doi.org/10.1162/jocn_a_01257

Thorat, S., Proklova, D., & Peelen, M. V. (2019). *The nature of the animacy organization in human ventral temporal cortex*. Retrieved from https://arxiv.org/abs/1904.02866v1

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

Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S.-M., & Carlson, T. A. (2016). Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *NeuroImage*, *132*, 59–70. https://doi.org/10.1016/j.neuroimage.2016.02.019

Watson, D. M., Young, A. W., & Andrews, T. J. (2016). Spatial properties of objects predict patterns of neural response in the ventral visual pathway. *NeuroImage*, *126*, 173–183. https://doi.org/10.1016/j.neuroimage.2015.11.043

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

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064. https://doi.org/10.3758/s13423-012-0295-x

Zachariou, V., Giacco, A. C. D., Ungerleider, L. G., & Yue, X. (2018). Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *Journal of Vision*, *18*(12), 3–3. https://doi.org/10.1167/18.12.3

Zellner, A., & Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics: Proceedings of the First InternationalMeeting* (pp. 585–603). Valencia: University of Valencia Press.