NeuroImage 221 (2020) 117139

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

# A humanness dimension to visual object coding in the brain

Erika W. Contini [a,c], Erin Goddard [a,b,c], Tijl Grootswagers [a,c,d], Mark Williams [c], Thomas Carlson [a,c,*]

[a] School of Psychology, The University of Sydney, Australia
[b] School of Psychology, UNSW Sydney, Australia
[c] Department of Cognitive Science, Macquarie University, Australia
[d] The MARCS Institute for Brain Behaviour and Development, Western Sydney University, Australia

A B S T R A C T

Neuroimaging studies investigating human object recognition have primarily focused on a relatively small number of object categories, in particular, faces, bodies, scenes, and vehicles. More recent studies have taken a broader focus, investigating hypothesized dichotomies, for example, animate versus inanimate, and continuous feature dimensions, such as biologically similarity. These studies typically have used stimuli that are identified as animate or inanimate, neglecting objects that may not fit into this dichotomy. We generated a novel stimulus set including standard objects and objects that blur the animate-inanimate dichotomy, for example, robots and toy animals. We used MEG time-series decoding to study the brain's emerging representation of these objects. Our analysis examined contemporary models of object coding such as dichotomous animacy, as well as several new higher order models that take into account an object's capacity for agency (i.e. its ability to move voluntarily) and capacity to experience the world. We show that early (0–200 ms) responses are predicted by the stimulus shape, assessed using a retinotopic model and shape similarity computed from human judgments. Thereafter, higher order models of agency/experience provided a better explanation of the brain's representation of the stimuli. Strikingly, a model of human similarity provided the best account for the brain's representation after an initial perceptual processing phase. Our findings provide evidence for a new dimension of object coding in the human brain – one that has a "human-centric" focus.

## 1. Introduction

Human object recognition is fast, efficient (Thorpe et al., 1996) – and fundamental to our interactions with the world. The ventral temporal cortex (VTC) is widely accepted as a key structure for visual object perception (Caramazza and Shelton, 1998; Haxby et al., 2001; Ishai et al., 1999; Mahon et al., 2007). One hypothesized organisational principle in human and primate VTC is the animate-inanimate dichotomy (Kiani et al., 2007; Kriegeskorte et al., 2008a,b; Pinsk et al., 2009). In support of this view, neuroimaging studies have shown subregions of the VTC with distinct response preferences, including a medial to lateral organization of animate and inanimate objects in the brain (Chao et al., 1999; Kanwisher et al., 1997; Konkle and Caramazza, 2013; Mahon et al., 2007; Taylor and Downing, 2011). It is also well known that specific regions within VTC respond preferentially to images from particular categories, including faces, animals, bodies (Downing et al., 2006; Downing et al., 2001; Haxby et al., 1994; Puce et al., 1996; Sergent et al., 1992), tools (Chao et al., 1999; Chao and Martin, 2000) and places (Epstein et al., 1999; Epstein and Kanwisher, 1998; Taylor and Downing, 2011).

An alternative approach to understanding object representations in the brain is to study how objects are represented in distributed patterns of brain activity (Haxby et al., 2001; Ishai et al., 1999). Using multivariate pattern analysis (MVPA) (for reviews see Grootswagers et al., 2017; Haynes, 2015; Pereira et al., 2009), researchers can study patterns of brain activity and test hypotheses about the neural representation of object categories (Kriegeskorte and Kievit, 2013; Kriegeskorte et al., 2008a,b). Using the MVPA framework, studies examining the relative similarity/dissimilarity of individual object representations in VTC have evidenced that objects may be represented along continuous dimensions in a multidimensional representation space. Animate subcategories have been argued to be coded along an axis of biologically similarity to humans (Connolly et al., 2012; Sha et al., 2015). This animacy continuum, however, does not provide a clear prediction for subcategory differentiation within the inanimate domain, nor for how the brain would represent objects that blur the animate-inanimate distinction (e.g., robots and animal toys). Early fMRI studies have shown that stick figures bodies and cartoon faces activate the extrastriate body area (EBA) and the fusiform face area (FFA) (Downing et al., 2001; Kanwisher et al., 1997; Tong et al., 2000), respectively. On the one hand, one would expect the EBA and FFA to respond to the figures because in a minimalist form they convey information about category membership. On the other hand, observers clearly know these figures are not alive. Moreover, it is also unclear whether a

* Corresponding author. School of Psychology, University of Sydney, 2006, NSW, Australia.
*E-mail addresses:* thomas.carlson@sydney.edu.au, thomas.shadowgrey@gmail.com (T. Carlson).

https://doi.org/10.1016/j.neuroimage.2020.117139
Received 4 October 2019; Received in revised form 27 May 2020; Accepted 2 July 2020
Available online 11 July 2020

continuum centred around 'animacy' best captures the dimension along which neural responses vary. Sha et al. (2015), for example, proposed that the neural representation of objects is better characterised according to the object's ability to perform goal-directed actions (see also Haxby et al., 2020; Thorat et al., 2019). Critically, there are many related factors to biologically similarity and agency that are known to influence human perception of objects (Gobbini et al., 2011; Gray et al., 2007). This raises the question about whether these factors also might be used as organisational principles for the brain's representation of objects.

In the present study, we used magnetoencephalography (MEG) to characterise the brain's neural representations of objects, and to explore their temporal dynamics. We studied the brain's emerging representation of 120 object stimuli and tested a wide range of models that might account for these representations using representational similarity analysis (RSA) (Kriegeskorte and Kievit, 2013; Kriegeskorte et al., 2008a,b). We found that, after an initial period of perceptual processing, higher order category models and models of agency and human-related experiences account for brain's representations of these objects. Notably, the model that best accounted for later stage representations of objects was a "human-centric" model, which describes objects in terms of their similarity to humans.

## 2. Materials and methods

### 2.1. Data Availability

Data from these experiments are freely available for download from the Open Science Framework (doi: https://doi.org/10.17605/OSF.IO/A52VT). This online repository includes deidentified raw data from the MEG and behavioural experiments, all stimuli used in the experiments, and the Matlab code used to perform all analyses and generate each figure in this manuscript.

### 2.2. Participants

Twenty-four English-speaking volunteers (18 female) with an average age of 24.7 years (SD = 5.47; range = 18–37) were recruited from the Macquarie University community. Informed written consent was obtained prior to participation, and participants were financially compensated for their time. All participants self-reported normal or corrected-to-normal vision (wearing of contacts was allowed), were free of medical conditions, and were not currently taking any neuroactive medications. This study was approved by the Macquarie University Human Research Ethics Committee.

### 2.3. Stimuli

Stimuli consisted of 120 naturalistic images of objects (Fig. 1), which were displayed on a uniform grey background. Twelve object categories were used in the study: six animate (humans, primates, domestic animals, birds, fish, invertebrates) and six inanimate (plants, robots, machines, tools, toys, other non-moving objects). In this stimulus set, animate is defined as living animals, in line with previous research (Caramazza and Shelton, 1998; Carlson et al., 2013; Connolly et al., 2012; Gobbini et al., 2011; Kriegeskorte et al., 2008a,b; Sha et al., 2015). Categories were selected to include ones similar to those used by Sha et al. (2015), with the addition of robots and toys to address the questions about agency and experience. We also included machines, which, like robots, had moving parts, but did not have the humanistic/animalistic/agentic properties. Stationary objects were also included, which neither moved nor had humanistic/animalistic/agentic properties.

### 2.4. MEG experimental procedure

For the experimental task, participants completed eight blocks of 398 trials (3184 trials in total). Within each block exemplars were presented for 100 ms, with a random inter-trial interval ranging between 750 and 1000 ms. The eight blocks were collected in a single session totalling approximately 1 h of MEG recording time. Stimuli were presented in a predetermined pseudo-randomised order, such that for each trial, the preceding image had an equal probability of being from any one of the 12 object categories. The ordering of the 8 blocks was pseudo-randomised across participants.


**Fig. 1.** Stimuli from each of the 12 object categories. Animate object categories are ordered vertically according to the biological classes animacy continuum (Sha et al., 2015). Brackets show two examples of different groupings of the stimuli: living vs. non-living and animate vs. inanimate.

Across trials, object images were manipulated in two ways to reduce the effects of low-level stimulus properties on our data. Firstly, a left-right flipped version of each image was included in the stimulus set, resulting in a total of 240 stimuli from 120 object images. Secondly, during image presentation, stimuli appeared in one of four locations while participants maintained fixation on a central marker, thus varying retinal location of the stimulus images. The four locations were defined by a shift from central presentation towards each of the four corners of the screen, where each stimulus location overlapped the central fixation point (details in *Display Apparatus* below). Each stimulus was presented three times at each location. This resulted in a total of 2880 trials (240 stimuli x 4 locations x 3 repetitions = 2880 trials). The additional trials were not included in the analysis: these included the first and last trial of each block, as well as 288 repeat trials that were added for the attention task (see below).

### 2.5. Attention task

During the experiment, participants completed a one-back attention task, where they were required to press a button whenever an object image was repeated consecutively. Participants received feedback about their accuracy on the task at the completion of each block. The mean accuracy across participants was 87.38% ($SD = 7.28\%$), with an average reaction time of 535 ms ($SD = 51$ ms). Due to a malfunction of the response button during the experiment, accuracy and reaction times were missing for one of our 24 participants, as well as for one out of the eight blocks for each of two further participants. These participants were still instructed to perform the task and were unaware that the button was not recording their responses.

### 2.6. Display Apparatus

Participants lay supine in the magnetically shielded recording room. Using an InFocus IN5108 projector situated outside the chamber, stimuli were projected onto a mirror, which reflected the image onto the ceiling, located approximately 113 cm above the participant. The total screen area was 20 × 15 degrees of visual angle (DVA). Throughout the experiment the screen background was held at a mean grey, and subjects were instructed to fixate on a black central fixation point (diameter of 0.1 DVA) that was always present. All stimulus locations were within a 6.9 DVA square, centred on the fixation point. Each stimulus consisted of a 256 × 256 pixel image (containing the segmented colour object) that was drawn to a 4.9 × 4.9 DVA square. Stimuli were presented one at a time, in one of four locations aligned with the upper left, upper right, lower left, or lower right corner of the 6.9 DVA square. A central square of 150 pixels (2.9 DVA) was common to all four stimulus locations. All stimuli were drawn as full colour segmented objects against a mean grey background (as in Fig. 1): the same mean grey as the screen outside the stimulus location. Upon stimulus presentation, a 50 × 50 pixel (1 × 1 DVA) white square simultaneously appeared in the bottom right corner of the projection, which was aligned with a photodetector attached to the mirror to accurately record the stimulus presentation time in the MEG recording. The experiment was run on a Dell PC desktop computer using MATLAB software (Natick, MA) and the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

### 2.7. MEG data acquisition

MEG data were recorded in the KIT-Macquarie Brain Research Laboratory using a 160-channel whole-head axial gradiometer (KIT, Kanazawa, Japan). Continuous data were acquired at a sampling rate of 1000 Hz, and were band-pass-filtered online from 0.03 to 200 Hz. MATLAB (2013b, Natick, MA) was used for all processing and statistical analyses of the data. Offline, we down-sampled the data to 200 Hz and epoched each trial into an event with a time window from −100 ms to 600 ms relative to stimulus onset. To reduce the dimensionality of the data, we applied Principal Components Analysis to the epoched data from the 160 gradiometers and retained the first *n* components that accounted for 99% of the variance. The number of components retained for each participant ranged from 14 to 72 (Mean = 34.21, SD = 18.90).

### 2.8. Classification analysis

For each participant, we used linear discriminant analysis to classify object/exemplar identity at the single trial level, training and testing classifiers on their ability to discriminate every possible exemplar pair of the 120 object images. We used cross-validated classification accuracy as a measure of how dissimilar the patterns of brain activity were for one exemplar compared to another (Nili et al., 2014). We did not attempt to model the effects of spatial position or left-right flip in our classification analysis, but instead used a single data label (the object identity) for data obtained from both the standard and left-right-flipped versions of the stimuli, as well as all four stimulus presentation locations. By including data from all variations of the stimuli, we sought to force the classifier to generalise beyond lower-level visual features, (such as the presence or absence of stimulation at a given location in the visual field), and instead use any neural correlate of object identity. These modifications to the stimulus presentation would have introduced extra noise into the signal across trials, so would tend to reduce classifier performance relative to unvarying stimuli, but they allowed us to better target higher-level object representations. For each time-point, we trained and tested a separate classifier to discriminate each pair of exemplar identities from the PCA components. We used a 10-fold cross-validation procedure, where the classifier was trained on data from 90% of the trials and then its accuracy was evaluated using its performance when classifying the remaining 10% of the data, so that the classifier was never tested on data that were included in the training set. This process was repeated 10 times, so that all trials were used as test data once each. D-prime (d') was used as the metric for classification accuracy.

### 2.9. Representational similarity analysis (RSA)

Classifier accuracies (d') were averaged across exemplar pairs to obtain the mean classifier performance for each time point. Additionally, to capture the pattern of classifier performance across exemplar pairs and compare this pattern with model predictions, we constructed a Representational Dissimilarity Matrix (RDM) for each time point. The RDM is a 120 × 120 matrix, symmetric along the diagonal, where each cell is the classification accuracy (d') for that pair of exemplars.

For each time point we compared each participants' observed neural RDM with model RDMs, where each model RDM was a 120 × 120 matrix derived from theory, computational modelling, or behavioural data (as described in detail below). This analysis, known as 'Representational Similarity Analysis' (RSA) (Kriegeskorte et al., 2008a,b) tests the relationship between models of interest and the group data, measuring how well the model RDMs account for the observed pattern of results. At each time point we used Kendall's tau-a to compute the rank order correlation between each candidate model and the neural data, then used these correlation values to compare candidate models in their ability to account for the neural data. Fig. 2 shows the model RDMs, which are described in detail below.

### 2.10. Low-level feature models (Fig. 2, models 1–3)

The HMAX and Jaccard silhouette models were included to test for the effects of low-level stimulus properties on the similarity/dissimilarity of neural responses, as measured using classifier performance.

*HMAX (model 1):* Computational model of low-level visual processes. We applied the HMAX model (Riesenhuber and Poggio, 1999; Serre et al., 2007) to simulate the responses of low-level visual areas. HMAX was applied to images at only a single image location and based on the standard orientation of each stimulus (i.e., not left-right flipped). The

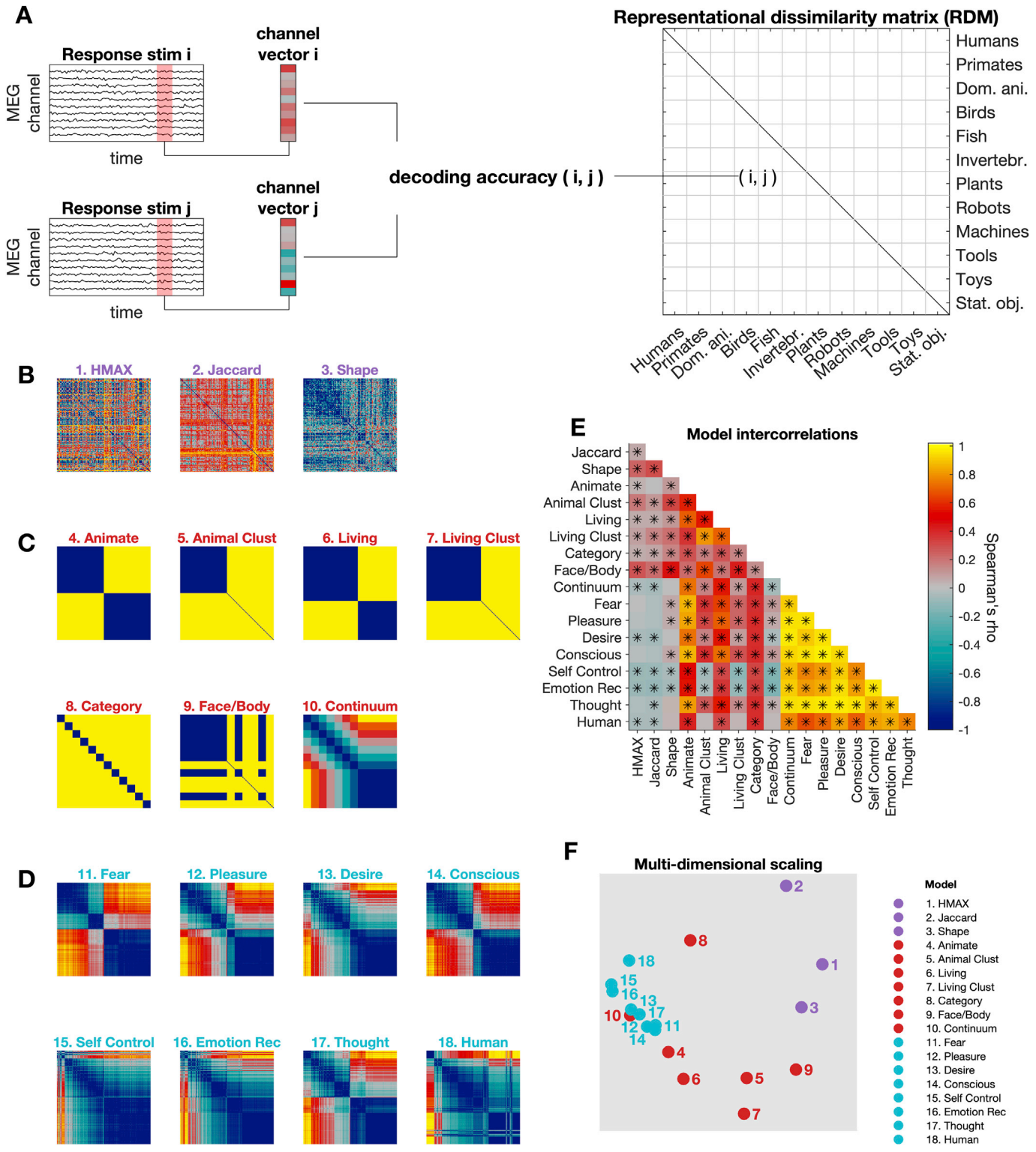


**Fig. 2.** Representational Similarity Analysis. (A). For all pairs of images, the dissimilarity between their MEG response patterns is stored in a Representational Dissimilarity Matrix (RDM). An RDM is created for each time point from the MEG data. (B–D) RDMs used for model testing. Model axes refer to all 120 image exemplars (grouped by category in the same order as Fig. 2A. Colour bar indicates predicted degree of dissimilarity between exemplar pairs. Models are grouped according to whether they are low-level feature models (B), contemporary models (C), or behavioural rating models (D). (E) Model correlation matrix. Cell colour indicates correlation strength (*p $<$ .05, adjusted for multiple comparisons across model pairs using a FDR of q $<$ 0.01), with yellow cells indicating a stronger correlation between models, while blue indicates a weak/no correlation. (F) MDS plot showing the representational geometry of model similarity in a 2-dimensional space. Models are colour coded according to whether they are low-level feature models, contemporary models, or behavioural rating models.

responses of the final HMAX layer (C2) for every stimulus were vectorized. We then generated the model RDM by taking the Euclidean distance between the vectorized model responses for each pair of stimuli.

*Jaccard (silhouette model; model 2):* An abstract shape model that measures the shape of each object in terms of the pixels that the image occupies (Jaccard, 1901). We generated the model RDM by comparing the overlapping silhouette regions of two images at a time and obtaining a measure of the difference. This model was generated based on the standard orientation of each stimulus (i.e., not flipped), independent of location.

*Shape-similarity model (model 3):* The shape similarity model was constructed using a visual search (Proklova et al., 2016, 2019). In short, shape dissimilarity was measured as the time it took participants to find a unique shape among identical distractor shapes in a visual search paradigm. For every pair of stimuli, we extracted the outline of the two objects and presented them on a 4 × 4 arrangement. The location of the oddball was randomly chosen on each trial. Participants responded whether the unique shape was on the left side or right side of the display using a key press (F and J for left and right). Each arrangement was preceded by a 500 ms fixation cross, and the arrangements remained on the screen until the participants responded. All combinations of stimuli were divided in 15 sets of 953 trials each, and 20 unique participants completed each set. A total of 300 Amazon's Mechanical Turk workers residing in either the United States of America or Canada, completed the experiment. The group had a mean accuracy of 0.8995 (s.d. 0.1673). Individual participant's reaction times were z-scored and then sign-flipped, and we then constructed the shape model by taking the median z-scored reaction times on correct trials for each pair of stimuli across participants.

### 2.11. Contemporary models of object representations (Fig. 2, models 4–10)

The contemporary models were created based on organisational structures proposed in previous studies, with the term 'contemporary' used to highlight that these reflect current theories of object category structure. Descriptions of each model are provided below.

*Dichotomy models (models 4 and 6):* The animate vs. inanimate dichotomy model (Caramazza and Shelton, 1998; Carlson et al., 2013; Cichy et al., 2014; Kriegeskorte et al., 2008a,b) is a category model that grouped all animate and inanimate objects separately (implying that objects within these groupings were more similar to each other, and more dissimilar to objects in the other grouping). Similarly, the living vs. non-living dichotomy model (Gainotti, 2000; Huth et al., 2012; Warrington and Shallice, 1984) grouped all living and non-living objects separately. The living category included the same items as the animate category but with the addition of plants.

*Cluster models (5 and 7):* The animal cluster model (model 5) is a single-category model that only grouped all animate objects together, suggesting that animate objects will be more similar to each other, and more dissimilar to all other objects, but that inanimate objects will not cluster. The living cluster model (model 7) follows the same principle, but grouping all living objects together. The cluster models were created to determine whether the effect of the dichotomy models was driven by cohesion within the in-group alone (i.e., animate, living), with more disparate object representations in the out-group category (i.e., inanimate, non-living) (Clarke and Tyler, 2014).

*Category model (model 8):* The category model was included as a measure of category individuation, as it proposes that items within individual categories have distinctly related patterns due to common visual and semantic properties, and these patterns are more different to those of objects from other categories (Clarke and Tyler, 2014). This model grouped each individual category as being more similar to within-category items and more dissimilar to other categories.

*Faces/bodies model (model 9):* Faces and bodies stand out as special categories for object recognition (Barragan-Jason et al., 2015; Cauchoix et al., 2014; Gobbini et al., 2011; Haxby et al., 2001; van de Nieuwenhuijzen et al., 2013) and so were of interest given the inclusion of toys and robots in our stimulus set. As such, the faces/bodies model is single-category model, grouping together all object categories that had faces or bodies, including all animate objects, as well as robots and toys.

*Continuum model (model 10):* The continuum model is a graded model based on the animacy continuum proposed by Sha et al. (2015). The continuum included a gradient of similarity between object categories that varied along a dimension related to biological classes, such that categories more similar to humans (biologically), would have more similar activity patterns, and those more dissimilar to humans would have activity patterns more similar to inanimate objects. For this model, plants were included on the continuum as they are a biological category and were represented on the continuum between invertebrates and inanimate objects. All non-living inanimate objects were treated as a single category, most dissimilar to the human category.

### 2.12. Behavioural-rating models for agency/experience models (Fig. 2, models 11–18)

The behavioural-rating models include the agency/experience models (models 11–17) and the human model (model 18). These models were created by obtaining behavioural ratings of the stimuli according to a specific question (detailed below). A total of 325 Amazon's Mechanical Turk workers residing in either the United States of America or Canada, completed one of the eight surveys online (number of participants per survey ranged from 40 to 43). Participants included 146 females (1 other, 1 no response), and had an average age of 35.27 years (SD = 10.26, range = 18.9–70.8; one age value missing). In each survey we asked workers to answer a single question for each of the stimuli:

11. *Fear* – How much is it capable of feeling afraid or fearful?
12. *Pleasure* – How much is it capable of experiencing physical or emotional pleasure?
13. *Desire* - How much is it capable of longing or hoping for things?
14. *Consciousness* - How much is it capable of having experiences and being aware of things?
15. *Thought* - How much is it capable of thinking?
16. *Emotion-recognition* - How much is it capable of understanding how others are feeling?
17. *Self-Control* - How much is it capable of exercising self-restraint over desires, emotions or impulses?
18. *Human* – How similar is this to a human?

Surveys for models 11–17 were based on a subset of the mental capacity surveys used in Gray et al. (2007), which vary as to how much they loaded onto the author's 'Experience' and 'Agency' factors that were established in their study. The seven agency/experience models were based on the results of these surveys. The 'Human' survey (18) was added to address a meta-representational idea of categorisation, that of "human-ness": a complex factor which may encompass biology, agency, and visual similarity. Each survey required participants to rate all 120 images on a 7-point scale from '*Not at all*' to '*Very much so*' in response to the specific question. Each survey took approximately 10 min to complete and participants were financially compensated for their time. The surveys were created and administered using the Qualtrics online survey platform. For each survey, participants provided voluntary consent and basic demographic information before completing the survey. Participants were only allowed to complete one of the eight surveys available, resulting in unique individuals for each survey. Stimulus order was randomised separately for each participant.

To construct the models based on agency and experience (shown in Fig. 2), a RDM was created for each set of survey responses by obtaining the absolute difference between image ratings for each pairwise comparison of the 120 images, using the mean ratings of each image. These RDMs, based on the survey ratings, provide hypothetical models of the

degree of dissimilarity between the neural responses associated with each image. For graphical purposes, we scaled these difference values between 0 and 1 for each model, such that warmer colours indicate greater dissimilarity, while cooler colours depict greater similarity between the neural representations in the pair-wise comparison.

### 2.13. Model intercorrelations (Fig. 2E)

As the models we used in this study were not orthogonal, we measured the degree of overlap by performing correlations (Spearman) between each of the models (see Fig. 2E). By evaluating the strength of these correlations, we obtained an estimate of how much the models overlap in terms of the hypotheses being tested. Of particular note, the behavioural-rating models based on the agency and experience factors from Gray et al. (2007) and the human model we created were all highly correlated (see clustering in Fig. 2F MDS plot of the representational geometry): this was not surprising as these models all capture slightly different aspects of similarity to humans.

In this study, we aimed to select stimuli that were visually diverse within each subcategory, to minimise the extent to which visual similarity would produce seemingly 'categorical' patterns of results. The model correlation data suggests that our stimulus set provided good separation of visual similarity and object category, since few models correlated with the visual feature models. Importantly, this should minimise the contribution of low-level visual similarity when we evaluate our hypothesis driven models. Exceptions to this included the animal cluster, category, and faces/bodies models, which each showed a significant correlation with one, or both of the HMAX and Jaccard models. This suggests that despite our stimulus diversity, there was still greater visual homogeneity of exemplars within the category groupings in these models than between category groupings. This means that, particularly for the animal cluster, category, and faces/bodies models, any correlation between these models and the observed pattern of classifier performance could be driven by low-level visual similarity rather than by the higher-level category structure represented by these models.

### 2.14. Statistical evaluation of models

For each participant, we computed the correlation between each model and the time varying RDMs constructed from their MEG data. The time varying model correlations (Fig. 3) were compared against zero using sign-rank tests, and the resulting p-values were FDR-corrected for time points at $q = 0.01$. For the time windowed analysis (Fig. 4), the model correlations were tested against each other at each time window using t-tests, and the resulting p-values were FDR-corrected (across all model comparisons) for at $q = 0.01$. The FDR correction was done separately for the three time windows.

## 3. Results

### 3.1. Decoding object exemplars from the MEG recordings

We scanned participants using MEG while they viewed 120 object stimuli and applied multivariate pattern analysis to the MEG sensor recordings at each time point, measuring how well the classifiers could decode the stimulus the participants were viewing. To study the brain's representation of the objects at each time point, we ran the decoding analysis for all possible pairwise combinations of the 120 object stimuli. These data were used to create time-varying RDMs identical in size to model RDMs (Fig. 2).

We first confirmed we could decode the objects from the MEG recordings. Fig. 3A shows the average performance of the classifier across all the pairwise combinations of objects. The results show sustained decoding of object exemplars from 50 ms post stimulus-onset to the end of the time window (600 ms) with peak decoding performance at approximately 105 ms post stimulus onset. These results are consistent with previous MEG decoding studies examining the emerging representation of objects in humans (Carlson et al., 2013; Cichy et al., 2014; Goddard et al., 2016).

### 3.2. The dynamic representation of objects

How does the brain's representation of objects unfold over time? Having established that we could decode the individual object images, we next tested a range of hypotheses about category representations by comparing the observed neural RDMs with the model RDMs at each time point, using RSA (Kriegeskorte and Kievit, 2013; Kriegeskorte et al., 2008a,b). The neural RDM at each time point describes the brain's representation of the stimuli at that time. The models (Fig. 2B–D) attempt to explain a proportion of the variance in this structure. Formally, the models were evaluated by computing the rank correlation between the neural RDMs and each model RDM (Fig. 3B and C).

### 3.3. Low-level models and shape similarity (Fig. 3B)

We evaluated two low-level models and one high level shape similarity model. The low-level models were designed to test how primitive visual features account for the brain's representation of the stimuli. The Jaccard (i.e., silhouette) model evaluates differences in the retinal projection of the stimuli (Jaccard, 1901). The HMAX model is based on a simulation of the response of early visual areas (Serre et al., 2007). The Jaccard and HMAX models both were significantly correlated with the neural RDMs during early stages in the time course, peaking at 75 and 105 ms respectively, and were no longer significant predictors after 250 ms. This is in agreement with the established literature about the time-course of visual object recognition, with responses related to lower-level visual stimulus properties occurring earlier on, and more abstract semantic and categorical responses occurring later (Carlson et al., 2014; Carlson et al., 2013; Cichy et al., 2014; Clarke and Tyler, 2014). Note that the models show high correlations even though our design incorporated left-right flips of the images and spatial displacement of the images to reduce the influence of low-level stimulus properties. The low-level models were generated using only the standard orientation of each stimulus at a fixed position, yet could still predict the data after these transformations, affirming the importance of low-level visual similarity in the initial representation of the stimuli. We additionally tested a high level shape similarity model based on human behaviour (Proklova et al., 2016, 2019). This model notably should not be as influenced by retinotopic factors (i.e. left-right flips of the images and spatial displacement), as it is based on observers' global perception shape. The shape similarity model peaked at 115 ms and was significant for much of the time period. This result affirms previous studies showing shape similarity provides a good account of the brain's early representation of the stimuli (Proklova et al., 2016, 2019). It is also notable that the shape similarity model peaked later than the two other low-level models, which accords with the view the higher order visual areas discount retinotopic differences in the emerging representation visual objects (DiCarlo et al., 2012).

### 3.4. Contemporary models: intermediate processing emphasizes faces and bodies (and shape similarity)

A wide range of theoretical models have been proposed to account for the brain's higher-order representation of objects. We tested how many of these models could account for the brain's emerging representation of the objects (Fig. 3B–D). The models we tested included a range of categorical models (e.g., animate versus inanimate), as well as a biological continuum model (Sha et al., 2015). We assessed their explanatory power using RSA and found that the models produced varying results. Of these models, starting at approximately 100 ms, the face/body category model had the most explanatory power. Notably, the shape model from human judgments followed a similar trajectory to

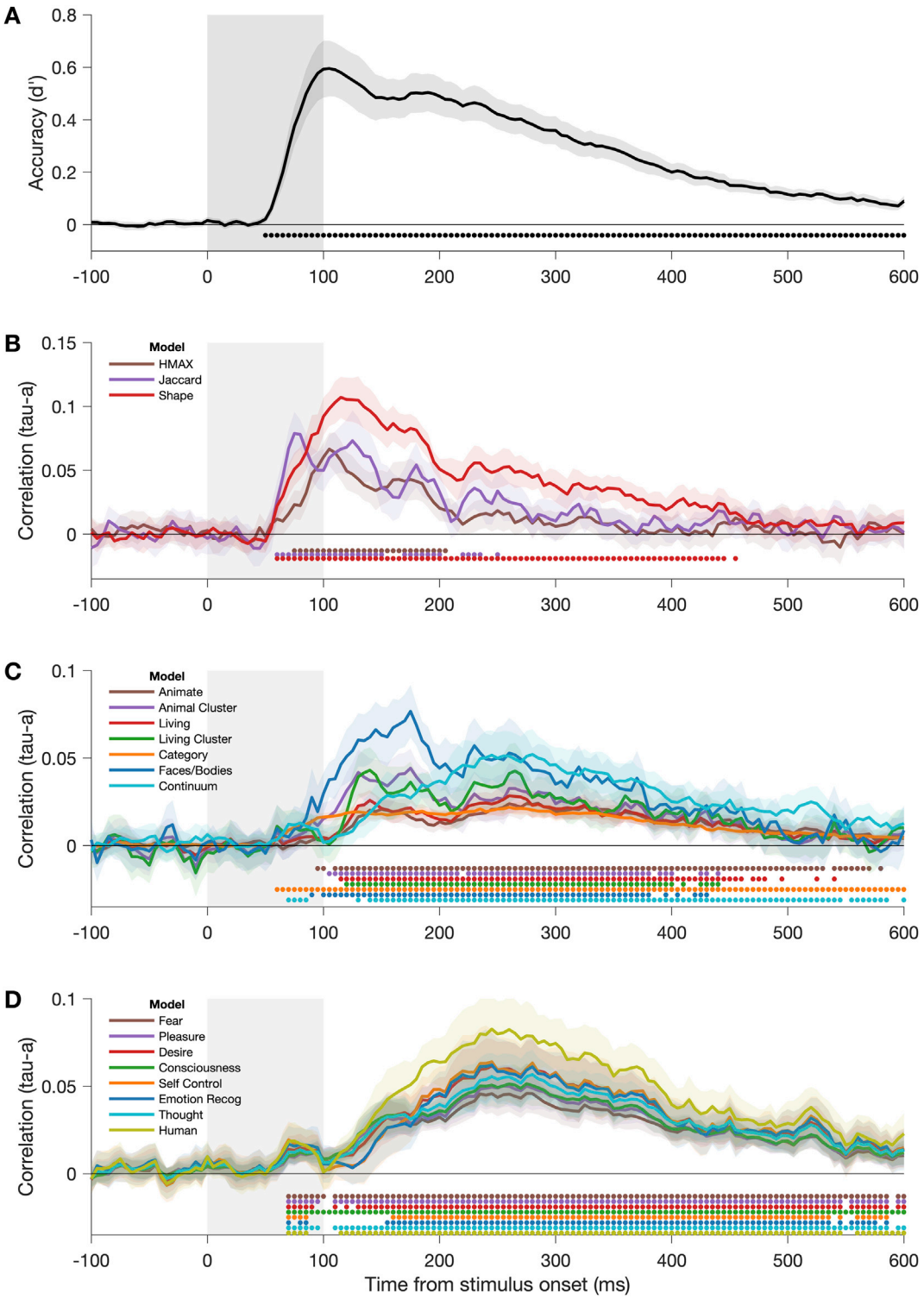


**Fig. 3.** Exemplar decoding and model testing. (A) Exemplar decoding. Average decoding performance (measured in d-prime) over time for all exemplar pairs. Grey bar indicates the period the image was on the screen. Error bars indicate 95% between-subject confidence intervals. Blue dots along the x-axis indicate time points at which decoding performance was significantly above chance (one-tailed *t*-test, adjusted for multiple comparisons across time points using a FDR of $q < 0.01$). (B–C) Model testing. Correlation between the classifier data and (B) the low-level visual feature models, (C) contemporary models, and (D) behavioural-rating models. Grey bar indicates the period the image was on the screen. Shaded area indicates the 95% confidence interval of the between-subject means. Colour coded dots along the x-axis indicate time points where the model provides a significant account for the data using a Kendall's Tau-a correlation (one-tailed *t*-test, adjusted for multiple comparisons across time points using an FDR of $q < 0.01$).

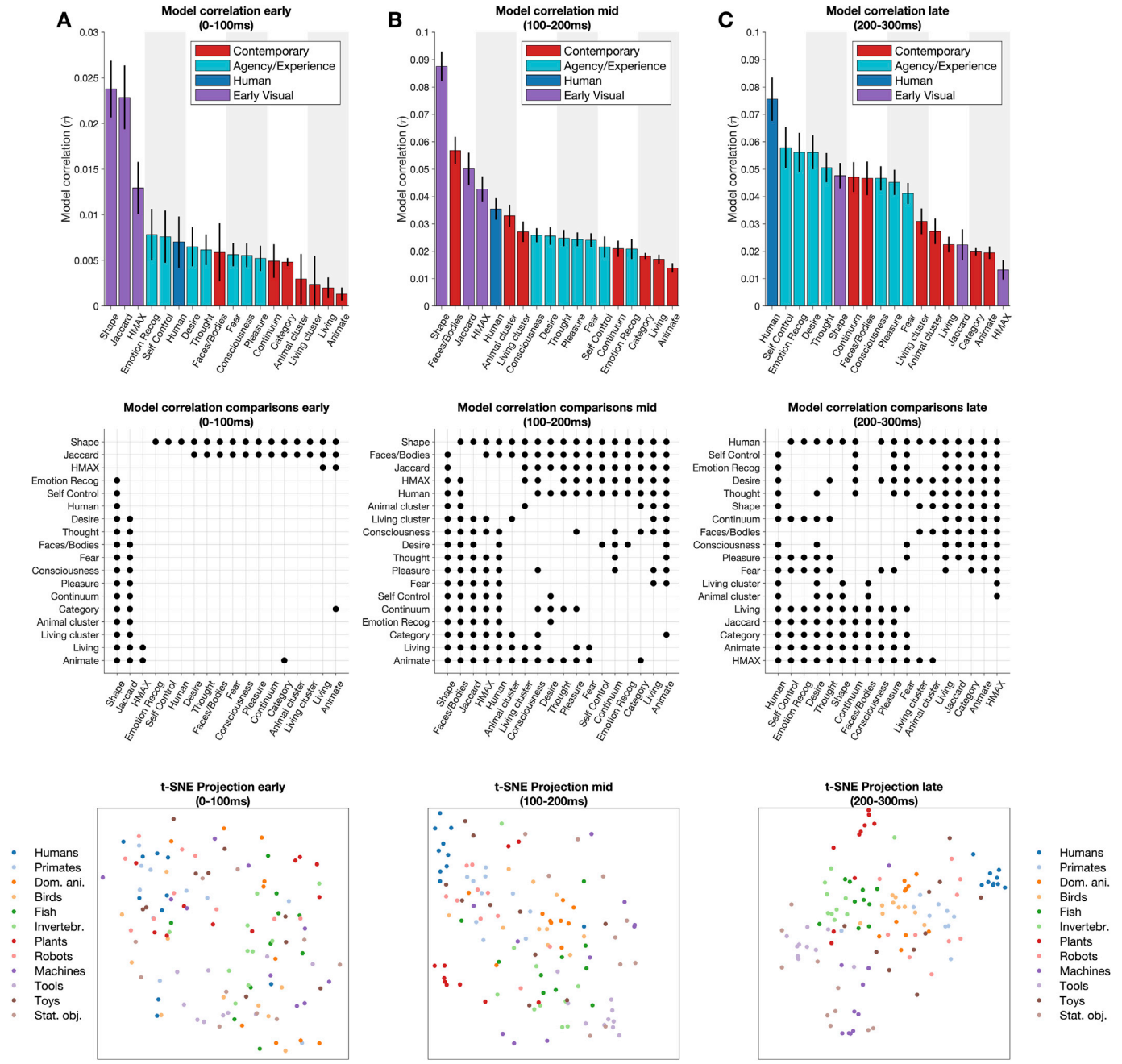


**Fig. 4.** Model correlations (Kendall's tau-a) in the early (0–100 ms; A), middle (100–200 ms; B) and late (200–300 ms; B) time windows. Models are arranged in order of highest average rank correlation within each time window (highest on the left), and error bars indicate 95% confidence interval of the between-subject means of the model correlation. The matrix below each plot shows significant differences between model correlations ($p < .05$, adjusted for model comparisons using an FDR of $q < 0.01$). Dots at intersections indicate a significant difference. Below each matrix, two-dimensional embeddings of the stimuli are shown. The embeddings were computed using t-SNE (Maaten and Hinton, 2008). The distance between two points in this embedding reflects their neural dissimilarity. Points are colour-coded according to category.

this model (compare Fig. 3B and C), which in part might be attributed to the high correlation between the two models (see Fig. 2E). At approximately 300 ms, there appeared to be a transition. Here, the faces/body model, which was the best performing model in between 100 and 200 ms, has declined and the biological continuum model increased its performance to have comparable explanatory power. Interestingly, the animacy model was among the weakest performing models, despite a number of studies showing animacy provides a significant account of the human and primate brain's higher-order representation of objects (e.g., Carlson et al., 2013; Cichy et al., 2014; Kiani et al., 2007; Kriegeskorte et al., 2008b).

### 3.5. Models of agency, experience, and human similarity

Higher order factors such as human similarity and agency are known to influence human perception of objects (Gobbini et al., 2011; Gray et al., 2007). To assess these attributes, we collected behavioural ratings for the stimuli about various higher order attributes (e.g., capacity to experience pleasure), to generate a new set of models. We then tested whether these models could account for the brain's emerging representation of objects (Fig. 3D). Across the models, the results were very similar, which can be attributed to the high level of overlap in their internal structure (see Fig. 2E). The models all show an initial peak at

approximately 90 ms, and then rise to a more significant peak at about 245–280 ms. This pattern was also observed for the biological continuum model, as seen in Fig. 3C. The best performing model in the later time window is the human similarity model, which is based on the question "How similar is this (object) to a human?"

### 3.6. Human-ness and agency/experience models account for late representations

Our analysis of the models' performance in the time series broadly indicated three distinct stages of processing. Early in the time series (<100 ms) the low-level feature models performed the best. Next, in the intermediate time from 100 ms to 200 ms, the face/body and shape similarity models peaked in their performance. Finally, the higher-order models based on agency and experience, and the biological continuum model, showed a slow rise that peaked about 270 ms. To quantify these observations, we discretised the data in three 100 ms time windows (0–100 ms; 100–200 ms; 200–300 ms) in which we compared all model performances. Fig. 4 shows the results of the windowed RSA analysis. The top row shows the models ordered by performance for each time period. Below each plot is matrix displaying significant differences in performance between models. The bottom row shows visualisations of the representations at each stage constructed by projecting the data into two dimensions using t-SNE (Maaten and Hinton, 2008).

To compare the models, we conducted a series of t-tests between all pairwise models to assess between-model performance in each of these time windows separately (adjusted for multiple comparisons across model comparisons at FDR of $q < 0.01$). In the early time window (0–100 ms), the high-level shape model had the highest correlation (Fig. 4A). This model outperformed all the other models excluding Jaccard and HMAX. In the second time window (100–200 ms), the shape model again outperformed all of the other models, including the face/body model, which it was highly correlated with (Fig. 2E). In the final time window (200–300 ms), the human similarity model was the best model overall (Fig. 4C), significantly better than all the other models excluding the face/body model. Overall, the results of the windowed analysis agree with the observations from the time series data. Shape similarity provided the best fit for early and intermediate stage representations. In late stage object processing, humanness was the best performing model, while the explanatory power of shape similarity diminished.

## 4. Discussion

Many classifications, such as animate/inanimate and living/non-living, have been proposed as organisational principles for the brain's representation of objects. Here we sought to provide an in-depth evaluation of contemporary models of visual object representations by evaluating their capacity to account for neural responses to a diverse range of object stimuli. In addition to these contemporary models and models based on low-level visual similarity, we created new theoretical and behaviour-based models. To test the predictive power of these models, we included novel stimuli that did not conform to the typical categories, such as robots and toys. Our results showed shape similarity best accounted for the brain's representation of the stimuli in early (<100 ms) and at intermediate stages (100–200 ms) processing. At later stages of processing, we found the brain employs a richer encoding scheme that incorporates an object's capacity for agency and experience. Of these models, the best performing was one based on the broad concept of human-similarity.

Our findings are consistent with accepted knowledge about the flow of information in human object recognition (for review see Contini et al., 2017). This multi-stage processes begins with processing low-level visual properties of the stimulus, presumably in early visual cortex. These early representations are then subsequently transformed into higher order representations incorporating category structure (Carlson et al., 2013; Cichy et al., 2016; Contini et al., 2017) and semantic information (Carlson et al., 2014; Clarke and Tyler, 2014). We found the best performing models at early time points were the low-level feature models (Shape, Jaccard, and HMAX), while higher order models (based on agency and experience) progressively fit the data better later in the time series.

One of the most striking results was that one of the lowest performing models was the animate vs. inanimate model, despite being a well-established model in the literature (Caramazza and Shelton, 1998; Carlson et al., 2013; Cichy et al., 2014; Kiani et al., 2007; Kriegeskorte et al., 2008a,b; Proklova et al., 2016). Our study included stimuli that do not clearly have membership in the animate or inanimate categories (Bracci and Op de Beeck, 2016; Carlson et al., 2013; Cichy et al., 2014; Konkle and Caramazza, 2013; Kriegeskorte et al., 2008a,b; Proklova et al., 2016). The poor performance of the animate vs. inanimate model (and similarly the living vs. non-living model) likely could be accounted for by the inclusion of robots and toys. For example, visually inspecting the t-SNE plots (Fig. 4) shows that robots are represented closer to humans and animate objects than to inanimate objects. This suggests that an animate/inanimate distinction is not the best way to classify these stimuli, and further highlights the impact of stimulus selection on defining the organisation of object categories (c.f. Carlson et al., 2018; Goddard et al., 2018). Indeed, a recent fMRI study by Bracci et al. (2017) showed that visually confusing objects (e.g., a mug in the shape of a cow) exhibited neural activity patterns that were more similar to animate objects (i.e., an actual cow) than inanimates (Bracci et al., 2017). Furthermore, as exemplar typicality affects the distinctiveness of category representations (Iordan et al., 2016), the inclusion of these ambiguous object categories may have disproportionately affected a strict dichotomous categorisation model.

The strong performance of the human model in late stage processing builds on our existing understanding of the representation of object categories in the brain. This aligns with recent work demonstrating the role of higher order constructs like agency in the visual representation of objects (Haxby et al., 2020); and extends on the continuum idea of Connolly et al. (2012) and Sha et al. (2015), as it represents a type of human-similarity continuum (see also Thorat et al., 2019). Unlike the animacy continuum that is based on biological classes, the human model was not limited by biology (Gobbini et al., 2011; Tong et al., 2000). Results from an fMRI study by Gobbini et al. (2011) are also consistent with a level of cross-over between animate/inanimate object categories that does not fit into this dichotomy, nor a continuum based on biological classes. The authors compared human observers' perception of human faces and robots and found that robots evoked activation in areas associated with faces (though to a lesser extent than humans), while also activating object areas and areas associated with mechanical movements. This supports the idea of more a complex model of object categorisation that incorporates factors such as agency and human-related experiences. Given the relative strength of our human-centric model in accounting for the data, the idea of "humanness" as an important dimension in the neural representation of objects warrants further exploration.

Our study tested a diverse range of stimuli and models. The richness of the dataset offers the future possibility of testing additional models. For example, one could envision a hybrid face model that indicates the presence of a face in the image with levels based on species: human, non-human primate, mammals, insects, inanimate objects with faces (toys, robots). The dataset accordingly will be made publicly available for future work. The diverse set of models tested has both limitations and benefits. One limitation is many of the models are not independent (Fig. 2E), which creates issues in interpretability. For example, the fits of the face/body model and the shape model to the MEG data had similar trajectories (Fig. 3B and C). The two models are also correlated (2 E), so it remains unknown whether the explanatory power of each of the models is based on the presence of face or body in the image or the shape of the figure in the image. Issues such as these will need to be disentangled in future work using a stimulus set designed to target these individual research questions. One strength of testing such a wide range of models is

each individual models' fit to the data gives insight into the brain's representations of the stimuli. If, for example, we had tested the human similarity model in the absence of the agency and experience models, the nature of the humanness model would be considerably more abstract. Here, we can see from the overlap between the agency/experience models (Fig. 2E) that the notion of humanness is captured well by factors like capacity for agency and experience.

Our human-centric model likely encompasses a complex set of features, including both visual and conceptual factors. This is evidenced by the overlap between the human-centric model and the other models (HMAX, Jaccard, Animate, the two cluster models, and the agency and experience models, see Fig. 2). We did not impose a definition or any criteria against which people should rate the objects when asked 'How similar is it to a human?' (with responses from this survey used to generate the human model). Accordingly, we do not know which features people were using to rate object 'humanness', raising an interesting area for further investigation. The brain likely makes use of both visual and semantic information for representing objects (Carlson et al., 2014; Clarke and Tyler, 2014; Coggan et al., 2016). Our data suggests that the semantic component of object representations incorporates information about concepts such as function, agency, and human experience. Indeed, a recent study by Connolly et al. (2012) showed an overlap between regions sensitive to the perceived threat of animals and those associated with social cognition, highlighting the importance of agent-related dimensions to object processing.

Presently, we still do not have a clear understanding of how different semantic concepts relate to object representations and category structure. A recent model attempts to explain the neural representation of object attempts using a multidimensional framework (Martin, 2016). In this paper, the author suggests that neural patterns associated with objects are formed from complex interactive circuits based on a range of systems throughout the brain, including those associated with action, perception and emotion. This idea shifts the focus away from models based on categories, with a view to a more holistic approach to object representations that considers interactions between various circuits throughout the brain. In this multidimensional framework, it is essential to recognize that no single feature or attribute could be able to fully explain the richness of the brain's (multidimensional) representation of objects (see e.g., Thorat, et al., 2019). Recent fMRI studies have sought to identify principle axes of object representations in the brain (e.g., Connolly, et al., 2012; Sha et al., 2015; Thorat et al., 2019). In the present study, we show that the human similarity provided a strong account of late stage processing, highlighting "humanness" as a key feature in the human brain's representation of objects that shapes our experience of the world.

## Author contributions

E.C, E.G, M.W, T.C designed the study.
E.C, T.G collected the data.
E.C, E.G, T.G conducted the analysis.
E.C, E.G, T.G, T.C interpreted the results.
E.C and T.C wrote the manuscript.
All authors reviewed the manuscript.

## CRediT authorship contribution statement

**Erika W. Contini:** Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing. **Erin Goddard:** Conceptualization, Formal analysis, Supervision, Visualization, Writing - review & editing. **Tijl Grootswagers:** Formal analysis, Visualization, Writing - review & editing. **Mark Williams:** Conceptualization, Writing - review & editing. **Thomas Carlson:** Conceptualization, Funding acquisition, Supervision, Writing - original draft, Writing - review & editing.

## 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.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.117139.

## References

Barragan-Jason, G., Cauchoix, M., Barbeau, E.J., 2015. The neural speed of familiar face recognition. Neuropsychologia 75, 390–401.
Bracci, S., Kalfas, I., Op de Beeck, H., 2017. The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. BioRxiv 228932.
Bracci, S., Op de Beeck, H., 2016. Dissociations and associations between shape and category representations in the two visual pathways. J. Neurosci. 36, 432–444.
Brainard, D.H., 1997. The psychophysics toolbox. Spatial Vis. 10, 433–436.
Caramazza, A., Shelton, J.R., 1998. Domain-specific knowledge systems in the brain the animate-inanimate distinction. J. Cognit. Neurosci. 10, 1–34.
Carlson, T.A., Goddard, E., Kaplan, D.M., Klein, C., Ritchie, J.B., 2018. Ghosts in machine learning for cognitive neuroscience: moving from data to theory. Neuroimage 180, 88–100.
Carlson, T.A., Simmons, R.A., Kriegeskorte, N., Slevc, L.R., 2014. The emergence of semantic meaning in the ventral temporal pathway. J. Cognit. Neurosci. 26, 120–131.
Carlson, T.A., Tovar, D.A., Alink, A., Kriegeskorte, N., 2013. Representational dynamics of object vision: the first 1000 ms. J. Vis. 13.
Cauchoix, M., Barragan-Jason, G., Serre, T., Barbeau, E.J., 2014. The neural dynamics of face detection in the wild revealed by MVPA. J. Neurosci. 34, 846–854.
Chao, L.L., Haxby, J.V., Martin, A., 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. 2, 913–919.
Chao, L.L., Martin, A., 2000. Representation of manipulable man-made objects in the dorsal stream. Neuroimage 12, 478–484.
Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., Oliva, A., 2016. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Sci. Rep. 6, 27755.
Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat. Neurosci. 17, 455–462.
Clarke, A., Tyler, L.K., 2014. Object-specific semantic coding in human perirhinal cortex. J. Neurosci. 34, 4766–4775.
Coggan, D.D., Baker, D.H., Andrews, T.J., 2016. The role of visual and semantic properties in the emergence of category-specific patterns of neural response in the human brain. eNeuro 3.
Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32, 2608–2618.
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.
DiCarlo, J.J., Zoccolan, D., Rust, N.C., 2012. How does the brain solve visual object recognition? Neuron 73, 415–434.
Downing, P.E., Chan, A.W., Peelen, M.V., Dodds, C.M., Kanwisher, N., 2006. Domain specificity in visual cortex. Cerebr. Cortex 16, 1453–1461.
Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473.
Epstein, R., Harris, A., Stanley, D., Kanwisher, N., 1999. The parahippocampal place area: recognition, navigation, or encoding? Neuron 23, 115–125.
Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392, 598–601.
Gainotti, G., 2000. What the locus of brain lesion tells us about the nature of the cognitive defect underlying category-specific disorders: a review. Cortex 36, 539–559.
Gobbini, M.I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., Guazzelli, M., Pietrini, P., 2011. Distinct neural systems involved in agency and animacy detection. J. Cognit. Neurosci. 23, 1911–1920.
Goddard, E., Carlson, T.A., Dermody, N., Woolgar, A., 2016. Representational dynamics of object recognition: feedforward and feedback information flows. Neuroimage 128, 385–397.
Goddard, E., Klein, C., Solomon, S.G., Hogendoorn, H., Carlson, T.A., 2018. Interpreting the dimensions of neural feature representations revealed by dimensionality reduction. Neuroimage 180, 41–67.
Gray, H.M., Gray, K., Wegner, D.M., 2007. Dimensions of mind perception. Science 315, 619.

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. J. Cognit. Neurosci. 29, 677–697.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.
Haxby, J.V., Gobbini, M.I., Nastase, S.A., 2020. Naturalistic stimuli reveal a dominant role for agentic action in visual representation. Neuroimage 116561.
Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.M., Pietrini, P., Grady, C.L., 1994. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J. Neurosci. 14, 6336–6353.
Haynes, J.D., 2015. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. Neuron 87, 257–270.
Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76, 1210–1224.
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.
Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. U. S. A. 96, 9379–9384.
Jaccard, P., 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société vaudoise des sciences naturelles 37.
Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.
Kiani, R., Esteky, H., Mirpour, K., Tanaka, K., 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J. Neurophysiol. 97, 4296–4309.
Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3? Perception 36, 14-14.
Konkle, T., Caramazza, A., 2013. Tripartite organization of the ventral stream by animacy and object size. J. Neurosci. 33, 10235–10242.
Kriegeskorte, N., Kievit, R.A., 2013. Representational geometry: integrating cognition, computation, and the brain. Trends Cognit. Sci. 17, 401–412.
Kriegeskorte, N., Mur, M., Bandettini, P., 2008a. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4.
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60, 1126–1141.
Maaten, L.v. d., Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.
Mahon, B.Z., Milleville, S.C., Negri, G.A., Rumiati, R.I., Caramazza, A., Martin, A., 2007. Action-related properties shape object representations in the ventral stream. Neuron 55, 507–520.
Martin, A., 2016. GRAPES-Grounding representations in action, perception, and emotion systems: how object properties and categories are represented in the human brain. Psychon. Bull. Rev. 23, 979–990.
Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A toolbox for representational similarity analysis. PLoS Comput. Biol. 10, e1003553.
Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vis. 10, 437–442.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45, S199–S209.
Pinsk, M.A., Arcaro, M., Weiner, K.S., Kalkus, J.F., Inati, S.J., Gross, C.G., Kastner, S., 2009. Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. J. Neurophysiol. 101, 2581–2600.
Proklova, D., Kaiser, D., Peelen, M.V., 2016. Disentangling representations of object shape and object category in human visual cortex: the animate-inanimate distinction. J. Cognit. Neurosci. 28, 680–692.
Proklova, D., Kaiser, D., Peelen, M.V., 2019. MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. Neuroimage 193, 167–177.
Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J. Neurosci. 16, 5205–5215.
Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nat. Neurosci. 2, 1019–1025.
Sergent, J., Ohta, S., MacDonald, B., 1992. Functional neuroanatomy of face and object processing. A positron emission tomography study. Brain 115 Pt 1, 15–36.
Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., Poggio, T., 2007. A quantitative theory of immediate visual recognition. Prog. Brain Res. 165, 33–56.
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. J. Cognit. Neurosci. 27, 665–678.
Taylor, J.C., Downing, P.E., 2011. Division of labor between lateral and ventral extrastriate representations of faces, bodies, and objects. J. Cognit. Neurosci. 23, 4122–4137.
Thorat, S., Proklova, D., Peelen, M.V., 2019. The Nature of the Animacy Organization in Human Ventral Temporal Cortex. arxiv.org arXiv:1904.02866.
Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381, 520–522.
Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., Kanwisher, N., 2000. Response properties of the human fusiform face area. Cogn. Neuropsychol. 17, 257–280.
van de Nieuwenhuijzen, M.E., Backus, A.R., Bahramisharif, A., Doeller, C.F., Jensen, O., van Gerven, M.A., 2013. MEG-based decoding of the spatiotemporal dynamics of visual category perception. Neuroimage 83, 1063–1073.
Warrington, E.K., Shallice, T., 1984. Category specific semantic impairments. Brain 107 (Pt 3), 829–854.