

**NeuroImage**  
**A humanness dimension to visual object coding in the brain**  
--Manuscript Draft--

<b>Manuscript Number:</b>	NIMG-19-2442R1
<b>Article Type:</b>	Full Length Article
<b>Section/Category:</b>	Systems and molecule neuroimaging
<b>Corresponding Author:</b>	Thomas Aaron Carlson, PhD University of Sydney Sydney, AUSTRALIA
<b>First Author:</b>	Erika W Contini, PhD
<b>Order of Authors:</b>	Erika W Contini, PhD Erin Goddard, PhD Tijl Grootswagers, PhD Mark Williams, PhD Thomas Aaron Carlson, PhD
<b>Abstract:</b>	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 (0-200ms) 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   **Title:** A humanness dimension to visual object coding in the brain

2   **Authors:** Erika W. Contini<sup>1,3</sup>, Erin Goddard<sup>1,2,3</sup>, Tijl Grootswagers<sup>1,3</sup>, Mark Williams<sup>3</sup>

3   and Thomas Carlson<sup>1,3</sup>

4

5   **Affiliations:**

6   1. School of Psychology, The University of Sydney, Australia

7   2. School of Psychology, UNSW Sydney, Australia

8   3. Department of Cognitive Science, Macquarie University, Australia

9

10   **Corresponding author:**

11   Thomas Carlson

12   School of Psychology

13   University of Sydney, 2006 NSW Australia

14

15

16

17    **Abstract**

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

35    **Introduction**

36    Human object recognition is fast, efficient (Thorpe, Fize, & Marlot, 1996) – and fundamental to our  
37    interactions with the world. The ventral temporal cortex (VTC) is widely accepted as a key  
38    structure for visual object perception (Caramazza & Shelton, 1998; Haxby, et al., 2001; Ishai,  
39    Ungerleider, Martin, Schouten, & Haxby, 1999; Mahon, et al., 2007). One hypothesized  
40    organisational principle in human and primate VTC is the animate-inanimate dichotomy (Kiani,  
41    Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte, Mur, Ruff, et al., 2008; Pinsk, et al., 2009). In  
42    support of this view, neuroimaging studies have shown subregions of the VTC with distinct  
43    response preferences, including a medial to lateral organization of animate and inanimate objects in  
44    the brain (Chao, Haxby, & Martin, 1999; Kanwisher, McDermott, & Chun, 1997; Konkle &  
45    Caramazza, 2013; Mahon, et al., 2007; Taylor & Downing, 2011). It is also well known that  
46    specific regions within VTC respond preferentially to images from particular categories, including  
47    faces, animals, bodies (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Downing, Jiang,  
48    Shuman, & Kanwisher, 2001; Haxby, et al., 1994; Puce, Allison, Asgari, Gore, & McCarthy, 1996;  
49    Sergent, Ohta, & MacDonald, 1992), tools (Chao, et al., 1999; Chao & Martin, 2000) and places  
50    (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998; Taylor & Downing,  
51    2011).

52       An alternative approach to understanding object representations in the brain is to study how  
53    objects are represented in distributed patterns of brain activity (Haxby, et al., 2001; Ishai, et al.,  
54    1999). Using multivariate pattern analysis (MVPA) (for reviews see Grootswagers, Wardle, &  
55    Carlson, 2017; Haynes, 2015; Pereira, Mitchell, & Botvinick, 2009), researchers can study patterns  
56    of brain activity and test hypotheses about the neural representation of object categories  
57    (Kriegeskorte & Kievit, 2013; Kriegeskorte, Mur, & Bandettini, 2008). Using the MVPA  
58    framework, studies examining the relative similarity/dissimilarity of individual object  
59    representations in VTC have evidenced that objects may be represented along continuous  
60    dimensions in a multidimensional representation space. Animate subcategories have been argued to

61 be coded along an axis of biologically similarity to humans (Connolly, et al., 2012; Sha, et al.,  
62 2015). This animacy continuum, however, does not provide a clear prediction for subcategory  
63 differentiation within the inanimate domain, nor for how the brain would represent objects that blur  
64 the animate-inanimate distinction (e.g., robots and animal toys). Early fMRI studies have shown  
65 that stick figures bodies and cartoon faces activate the extrastriate body area (EBA) and the  
66 fusiform face area (FFA) (Downing, et al., 2001; Kanwisher, et al., 1997; Tong, Nakayama,  
67 Moscovitch, Weinrib, & Kanwisher, 2000), respectively. On the one hand, one would expect the  
68 EBA and FFA to respond to the figures because in a minimalist form they convey information  
69 about category membership. On the other hand, observers clearly know these figures are not alive.  
70 Moreover, it is also unclear whether a continuum centred around ‘animacy’ best captures the  
71 dimension along which neural responses vary. Sha et al. (2015), for example, proposed that the  
72 neural representation of objects is better characterised according to the object’s ability to perform  
73 goal-directed actions (see also Haxby, Gobbini, & Nastase, 2020; Thorat, Proklova, & Peelen,  
74 2019). Critically, there are many related factors to biologically similarity and agency that are known  
75 to influence human perception of objects (Gobbini, et al., 2011; Gray, Gray, & Wegner, 2007). This  
76 raises the question about whether these factors also might be used as organisational principles for  
77 the brain’s representation of objects.

78 In the present study, we used magnetoencephalography (MEG) to characterise the brain’s  
79 neural representations of objects, and to explore their temporal dynamics. We studied the brain’s  
80 emerging representation of 120 object stimuli and tested a wide range of models that might account  
81 for these representations using representational similarity analysis (RSA) (Kriegeskorte & Kievit,  
82 2013; Kriegeskorte, Mur, & Bandettini, 2008). We found that, after an initial period of perceptual  
83 processing, higher order category models and models of agency and human-related experiences  
84 account for brain’s representations of these objects. Notably, the model that best accounted for later  
85 stage representations of objects was a “human-centric” model, which describes objects in terms of  
86 their similarity to humans.

87 **Materials and Methods**

88 ***Participants***

89 Twenty-four English-speaking volunteers (18 female) with an average age of 24.7 years (SD =  
90 5.47; range = 18-37) were recruited from the Macquarie University community. Informed written  
91 consent was obtained prior to participation, and participants were financially compensated for their  
92 time. All participants self-reported normal or corrected-to-normal vision (wearing of contacts was  
93 allowed), were free of medical conditions, and were not currently taking any neuroactive  
94 medications. This study was approved by the Macquarie University Human Research Ethics  
95 Committee.

96

97 ***Stimuli***

98 Stimuli consisted of 120 naturalistic images of objects (Figure 1), which were displayed on a  
99 uniform grey background. Twelve object categories were used in the study: six animate (humans,  
100 primates, domestic animals, birds, fish, invertebrates) and six inanimate (plants, robots, machines,  
101 tools, toys, other non-moving objects). In this stimulus set, animate is defined as living animals, in  
102 line with previous research (Caramazza & Shelton, 1998; Carlson, Tovar, Alink, & Kriegeskorte,  
103 2013; Connolly, et al., 2012; Gobbini, et al., 2011; Kriegeskorte, Mur, Ruff, et al., 2008; Sha, et al.,  
104 2015). Categories were selected to include ones similar to those used by Sha et al. (2015), with the  
105 addition of robots and toys to address the questions about agency and experience. We also included  
106 machines, which, like robots, had moving parts, but did not have the humanistic/animalistic/agentic  
107 properties. Stationary objects were also included, which neither moved nor had  
108 humanistic/animalistic/agentic properties.

109

110 **[Insert Figure 1 here]**

111 ***MEG Experimental Procedure***

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

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

130

131 ***Attention Task***

132 During the experiment, participants completed a one-back attention task, where they were required  
133 to press a button whenever an object image was repeated consecutively. Participants received  
134 feedback about their accuracy on the task at the completion of each block. The mean accuracy  
135 across participants was 87.38% ( $SD = 7.28\%$ ), with an average reaction time of 535 ms ( $SD = 51$   
136 ms). Due to a malfunction of the response button during the experiment, accuracy and reaction

137 times were missing for one of our 24 participants, as well as for one out of the eight blocks for each  
138 of two further participants. These participants were still instructed to perform the task and were  
139 unaware that the button was not recording their responses.

140

141 ***Display Apparatus***

142 Participants lay supine in the magnetically shielded recording room. Using an InFocus IN5108  
143 projector situated outside the chamber, stimuli were projected onto a mirror, which reflected the  
144 image onto the ceiling, located approximately 113 cm above the participant. The total screen area  
145 was 20x15 degrees of visual angle (DVA). Throughout the experiment the screen background was  
146 held at a mean grey, and subjects were instructed to fixate on a black central fixation point  
147 (diameter of 0.1 DVA) that was always present. All stimulus locations were within a 6.9 DVA  
148 square, centred on the fixation point. Each stimulus consisted of a 256x256 pixel image (containing  
149 the segmented colour object) that was drawn to a 4.9x4.9 DVA square. Stimuli were presented one  
150 at a time, in one of four locations aligned with the upper left, upper right, lower left, or lower right  
151 corner of the 6.9 DVA square. A central square of 150 pixels (2.9 DVA) was common to all four  
152 stimulus locations. All stimuli were drawn as full colour segmented objects against a mean grey  
153 background (as in Figure 1): the same mean grey as the screen outside the stimulus location. Upon  
154 stimulus presentation, a 50x50 pixel (1x1 DVA) white square simultaneously appeared in the  
155 bottom right corner of the projection, which was aligned with a photodetector attached to the mirror  
156 to accurately record the stimulus presentation time in the MEG recording. The experiment was run  
157 on a Dell PC desktop computer using MATLAB software (Natick, MA) and the Psychophysics

158 Toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

159

160 **MEG Data Acquisition** MEG data were recorded in the KIT-Macquarie Brain Research Laboratory

161 using a 160-channel whole-head axial gradiometer (KIT, Kanazawa, Japan). Continuous data were

162 acquired at a sampling rate of 1000 Hz, and were band-pass-filtered online from 0.03 to 200 Hz.

163 MATLAB (2013b, Natick, MA) was used for all processing and statistical analyses of the data.

164 Offline, we down-sampled the data to 200Hz and epoched each trial into an event with a time

165 window from -100 ms to 600 ms relative to stimulus onset. To reduce the dimensionality of the

166 data, we applied Principal Components Analysis to the epoched data from the 160 gradiometers and

167 retained the first  $n$  components that accounted for 99% of the variance. The number of components

168 retained for each participant ranged from 14 to 72 (Mean = 34.21, SD = 18.90).

169

170 [Insert Figure 2 here]

171

172 **Classification analysis** For each participant, we used linear discriminant analysis to classify

173 object/exemplar identity at the single trial level, training and testing classifiers on their ability to

174 discriminate every possible exemplar pair of the 120 object images. We used cross-validated

175 classification accuracy as a measure of how dissimilar the patterns of brain activity were for one

176 exemplar compared to another (Nili, et al., 2014). We did not attempt to model the effects of spatial

177 position or left-right flip in our classification analysis, but instead used a single data label (the

178 object identity) for data obtained from both the standard and left-right-flipped versions of the

179 stimuli, as well as all four stimulus presentation locations. By including data from all variations of

180 the stimuli, we sought to force the classifier to generalise beyond lower-level visual features, (such

181 as the presence or absence of stimulation at a given location in the visual field), and instead use any

182 neural correlate of object identity. These modifications to the stimulus presentation would have

183 introduced extra noise into the signal across trials, so would tend to reduce classifier performance

184 relative to unvarying stimuli, but they allowed us to better target higher-level object representations.  
185 For each time-point, we trained and tested a separate classifier to discriminate each pair of exemplar  
186 identities from the PCA components. We used a 10-fold cross-validation procedure, where the  
187 classifier was trained on data from 90% of the trials and then its accuracy was evaluated using its  
188 performance when classifying the remaining 10% of the data, so that the classifier was never tested  
189 on data that were included in the training set. This process was repeated 10 times, so that all trials  
190 were used as test data once each. D-prime ( $d'$ ) was used as the metric for classification accuracy.

191

192 ***Representational Similarity Analysis (RSA)*** Classifier accuracies ( $d'$ ) were averaged across  
193 exemplar pairs to obtain the mean classifier performance for each time point. Additionally, to  
194 capture the pattern of classifier performance across exemplar pairs and compare this pattern with  
195 model predictions, we constructed a Representational Dissimilarity Matrix (RDM) for each time  
196 point. The RDM is a 120x120 matrix, symmetric along the diagonal, where each cell is the  
197 classification accuracy ( $d'$ ) for that pair of exemplars.

198 For each time point we compared each participants' observed neural RDM with model  
199 RDMs, where each model RDM was a 120x120 matrix derived from theory, computational  
200 modelling, or behavioural data (as described in detail below). This analysis, known as  
201 'Representational Similarity Analysis' (RSA) (Kriegeskorte, Mur, & Bandettini, 2008) tests the  
202 relationship between models of interest and the group data, measuring how well the model RDMs  
203 account for the observed pattern of results. At each time point we used Kendall's tau-a to compute  
204 the rank order correlation between each candidate model and the neural data, then used these  
205 correlation values to compare candidate models in their ability to account for the neural data. Figure  
206 2 shows the model RDMs, which are described in detail below.

207

208 ***Low-level feature models (Figure 2, models 1-3)***

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

212

213 *HMAX (model 1):* Computational model of low-level visual processes. We applied the HMAX  
214 model (Riesenhuber & Poggio, 1999; Serre, et al., 2007) to simulate the responses of low-level  
215 visual areas. HMAX was applied to images at only a single image location and based on the  
216 standard orientation of each stimulus (i.e., not left-right flipped). The responses of the final HMAX  
217 layer (C2) for every stimulus were vectorized. We then generated the model RDM by taking the  
218 Euclidean distance between the vectorized model responses for each pair of stimuli.

219

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

225

226 *Shape-similarity model (model 3):* The shape similarity model was constructed using a visual search  
227 (Proklova, Kaiser, & Peelen, 2016, 2019). In short, shape dissimilarity was measured as the time it  
228 took participants to find a unique shape among identical distractor shapes in a visual search  
229 paradigm. For every pair of stimuli, we extracted the outline of the two objects and presented them  
230 on a 4x4 arrangement. The location of the oddball was randomly chosen on each trial. Participants  
231 responded whether the unique shape was on the left side or right side of the display using a key  
232 press (F and J for left and right). Each arrangement was preceded by a 500ms fixation cross, and the  
233 arrangements remained on the screen until the participants responded. All combinations of stimuli  
234 were divided in 15 sets of 953 trials each, and 20 unique participants completed each set. A total of

235 300 Amazon's Mechanical Turk workers residing in either the United States of America or Canada,  
236 completed the experiment. The group had a mean accuracy of 0.8995 (s.d. 0.1673). Individual  
237 participant's reaction times were z-scored and then sign-flipped, and we then constructed the shape  
238 model by taking the median z-scored reaction times on correct trials for each pair of stimuli across  
239 participants.

240

241 ***Contemporary models of object representations (Figure 2, models 4-10)***

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

245

246 ***Dichotomy models (models 4 and 6):*** The animate vs. inanimate dichotomy model (Caramazza &  
247 Shelton, 1998; Carlson, et al., 2013; Cichy, Pantazis, & Oliva, 2014; Kriegeskorte, Mur, Ruff, et al.,  
248 2008) is a category model that grouped all animate and inanimate objects separately (implying that  
249 objects within these groupings were more similar to each other, and more dissimilar to objects in  
250 the other grouping). Similarly, the living vs. non-living dichotomy model (Gainotti, 2000; Huth,  
251 Nishimoto, Vu, & Gallant, 2012; Warrington & Shallice, 1984) grouped all living and non-living  
252 objects separately. The living category included the same items as the animate category but with the  
253 addition of plants.

254

255 ***Cluster models (5 and 7):*** The animal cluster model (model 5) is a single-category model that only  
256 grouped all animate objects together, suggesting that animate objects will be more similar to each  
257 other, and more dissimilar to all other objects, but that inanimate objects will not cluster. The living  
258 cluster model (model 7) follows the same principle, but grouping all living objects together. The  
259 cluster models were created to determine whether the effect of the dichotomy models was driven by

260 cohesion within the in-group alone (i.e., animate, living), with more disparate object representations  
261 in the out-group category (i.e., inanimate, non-living) (Clarke & Tyler, 2014).

262

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

268

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

275

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

284

285 ***Behavioural-rating models for agency/experience models (Figure 2, models 11-18)***

286 The behavioural-rating models include the agency/experience models (models 11 – 17) and the  
287 human model (model 18). These models were created by obtaining behavioural ratings of the  
288 stimuli according to a specific question (detailed below). A total of 325 Amazon's Mechanical Turk  
289 workers residing in either the United States of America or Canada, completed one of the eight  
290 surveys online (number of participants per survey ranged from 40 – 43). Participants included 146  
291 females (1 other, 1 no response), and had an average age of 35.27 years (SD = 10.26, range = 18.9 –  
292 70.8; one age value missing). In each survey we asked workers to answer a single question for each  
293 of the stimuli:

294

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

304

305 Surveys for models 11-17 were based on a subset of the mental capacity surveys used in Gray et al.  
306 (2007), which vary as to how much they loaded onto the author's 'Experience' and 'Agency'  
307 factors that were established in their study. The seven agency/experience models were based on the  
308 results of these surveys. The 'Human' survey (18) was added to address a meta-representational  
309 idea of categorization, that of "human-ness": a complex factor which may encompass biology,  
310 agency, and visual similarity. Each survey required participants to rate all 120 images on a 7-point  
311 scale from 'Not at all' to 'Very much so' in response to the specific question. Each survey took

312 approximately 10 minutes to complete and participants were financially compensated for their time.  
313 The surveys were created and administered using the Qualtrics online survey platform. For each  
314 survey, participants provided voluntary consent and basic demographic information before  
315 completing the survey. Participants were only allowed to complete one of the eight surveys  
316 available, resulting in unique individuals for each survey. Stimulus order was randomised  
317 separately for each participant.

318

319 To construct the models based on agency and experience (shown in Figure 2), a RDM was created  
320 for each set of survey responses by obtaining the absolute difference between image ratings for each  
321 pairwise comparison of the 120 images, using the mean ratings of each image. These RDMs, based  
322 on the survey ratings, provide hypothetical models of the degree of dissimilarity between the neural  
323 responses associated with each image. For graphical purposes, we scaled these difference values  
324 between 0 and 1 for each model, such that warmer colours indicate greater dissimilarity, while  
325 cooler colours depict greater similarity between the neural representations in the pair-wise  
326 comparison.

327

### 328 ***Model intercorrelations (Figure 2E)***

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

336 In this study, we aimed to select stimuli that were visually diverse within each subcategory, to  
337 minimise the extent to which visual similarity would produce seemingly ‘categorical’ patterns of

338 results. The model correlation data suggests that our stimulus set provided good separation of visual  
339 similarity and object category, since few models correlated with the visual feature models.  
340 Importantly, this should minimise the contribution of low-level visual similarity when we evaluate  
341 our hypothesis driven models. Exceptions to this included the animal cluster, category, and  
342 faces/bodies models, which each showed a significant correlation with one, or both of the HMAX  
343 and Jaccard models. This suggests that despite our stimulus diversity, there was still greater visual  
344 homogeneity of exemplars within the category groupings in these models than between category  
345 groupings. This means that, particularly for the animal cluster, category, and faces/bodies models,  
346 any correlation between these models and the observed pattern of classifier performance could be  
347 driven by low-level visual similarity rather than by the higher-level category structure represented  
348 by these models.

349

### 350 ***Statistical evaluation of models***

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

358

## 359 **Results**

360

### 361 **Decoding object exemplars from the MEG recordings**

362 We scanned participants using MEG while they viewed 120 object stimuli and applied multivariate  
363 pattern analysis to the MEG sensor recordings at each time point, measuring how well the

364 classifiers could decode the stimulus the participants were viewing. To study the brain's  
365 representation of the objects at each time point, we ran the decoding analysis for all possible  
366 pairwise combinations of the 120 object stimuli. These data were used to create time-varying RDMs  
367 identical in size to model RDMs (Figure 2).

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

375

### 376 **The dynamic representation of objects**

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

385

### 386 **Low-level models and shape similarity (Figure 3B).**

387 We evaluated two low-level models and one high level shape similarity model. The low-level  
388 models were designed to test how primitive visual features account for the brain's representation of  
389 the stimuli. The Jaccard (i.e., silhouette) model evaluates differences in the retinal projection of the

390 stimuli (Jaccard, 1901). The HMAX model is based on a simulation of the response of early visual  
391 areas (Serre et al., 2007). The Jaccard and HMAX models both were significantly correlated with  
392 the neural RDMs during early stages in the time course, peaking at 75 and 105 ms respectively, and  
393 were no longer significant predictors after 250 ms. This is in agreement with the established  
394 literature about the time-course of visual object recognition, with responses related to lower-level  
395 visual stimulus properties occurring earlier on, and more abstract semantic and categorical  
396 responses occurring later (Carlson, Simmons, Kriegeskorte, & Slevc, 2014; Carlson, et al., 2013;  
397 Cichy, et al., 2014; Clarke & Tyler, 2014). Note that the models show high correlations even though  
398 our design incorporated left-right flips of the images and spatial displacement of the images to  
399 reduce the influence of low-level stimulus properties. The low-level models were generated using  
400 only the standard orientation of each stimulus at a fixed position, yet could still predict the data  
401 after these transformations, affirming the importance of low-level visual similarity in the initial  
402 representation of the stimuli. We additionally tested a high level shape similarity model based on  
403 human behaviour (Proklova, et al., 2016, 2019). This model notably should not be as influenced by  
404 retinotopic factors (i.e. left-right flips of the images and spatial displacement), as it is based on  
405 observers' global perception shape. The shape similarity model peaked at 115ms and was  
406 significant for much of the time period. This result affirms previous studies showing shape  
407 similarity provides a good account of the brain's early representation of the stimuli (Proklova, et al.,  
408 2016, 2019). It is also notable that the shape similarity model peaked later than the two other low-  
409 level models, which accords with the view the higher order visual areas discount retinotopic  
410 differences in the emerging representation visual objects (DiCarlo, Zoccolan, & Rust, 2012).

411

412 **Contemporary models: Intermediate processing emphasizes faces and bodies (and shape  
413 similarity)**

414 A wide range of theoretical models have been proposed to account for the brain's higher-order  
415 representation of objects. We tested how many of these models could account for the brain's

416 emerging representation of the objects (Figure 3B-D). The models we tested included a range of  
417 categorical models (e.g., animate versus inanimate), as well as a biological continuum model (Sha  
418 et al., 2015). We assessed their explanatory power using RSA and found that the models produced  
419 varying results. Of these models, starting at approximately 100ms, the face/body category model  
420 had the most explanatory power. Notably, the shape model from human judgments followed a  
421 similar trajectory to this model (compare Figure 3B and Figure 3C), which in part might be  
422 attributed to the high correlation between the two models (see Figure 2E). At approximately 300ms,  
423 there appeared to be a transition. Here, the faces/body model, which was the best performing model  
424 in between 100-200ms, has declined and the biological continuum model increased its performance  
425 to have comparable explanatory power. Interestingly, the animacy model was among the weakest  
426 performing models, despite a number of studies showing animacy provides a significant account of  
427 the human and primate brain's higher-order representation of objects (e.g., Carlson et al., 2013;  
428 Cichy et al., 2014; Kiani et al., 2007; Kriegeskorte et al., 2008b).

429

### 430 **Models of agency, experience, and human similarity.**

431 Higher order factors such as human similarity and agency are known to influence human perception  
432 of objects (Gobbini et al., 2011; Gray et al., 2007). To assess these attributes, we collected  
433 behavioural ratings for the stimuli about various higher order attributes (e.g., capacity to experience  
434 pleasure), to generate a new set of models. We then tested whether these models could account for  
435 the brain's emerging representation of objects (Figure 3D). Across the models, the results were very  
436 similar, which can be attributed to the high level of overlap in their internal structure (see Figure  
437 2E). The models all show an initial peak at approximately 90ms, and then rise to a more significant  
438 peak at about 245 to 280 ms. This pattern was also observed for the biological continuum model, as  
439 seen in Figure 3C. The best performing model in the later time window is the human similarity  
440 model, which is based on the question "How similar is this (object) to a human?"

441

442 **Human-ness and Agency/Experience models account for late representations**

443 Our analysis of the models' performance in the time series broadly indicated three distinct stages of  
444 processing. Early in the time series (<100ms) the low-level feature models performed the best.  
445 Next, in the intermediate time from 100ms to 200ms, the face/body and shape similarity models  
446 peaked in their performance. Finally, the higher-order models based on agency and experience, and  
447 the biological continuum model, showed a slow rise that peaked about 270ms. To quantify these  
448 observations, we discretised the data in three 100ms time windows (0-100ms; 100-200ms; 200-  
449 300ms) in which we compared all model performances. Figure 4 shows the results of the windowed  
450 RSA analysis. The top row shows the models ordered by performance for each time period. Below  
451 each plot is matrix displaying significant differences in performance between models. The bottom  
452 row shows visualisations of the representations at each stage constructed by projecting the data into  
453 two dimensions using t-SNE (Maaten & Hinton, 2008).

454

455 To compare the models, we conducted a series of t-tests between all pairwise models to assess  
456 between-model performance in each of these time windows separately (adjusted for multiple  
457 comparisons across model comparisons at FDR of  $q < .01$ ). In the early time window (0-100ms),  
458 the high-level shape model had the highest correlation (Figure 4A). This model outperformed all the  
459 other models excluding Jaccard and HMAX. In the second time window (100-200ms), the shape  
460 model again outperformed all of the other models, including the face/body model, which it was  
461 highly correlated with (Figure 2E). In the final time window (200-300ms), the human similarity  
462 model was the best model overall (Figure 4C), significantly better than all the other models  
463 excluding the face/body model. Overall, the results of the windowed analysis agree with the  
464 observations from the time series data. Shape similarity provided the best fit for early and  
465 intermediate stage representations. In late stage object processing, humanness was the best  
466 performing model, while the explanatory power of shape similarity diminished.

467

468

469

470

471

472 **Discussion**

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

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

493 One of the most striking results was that one of the lowest performing models was the  
494 animate vs. inanimate model, despite being a well-established model in the literature (Caramazza &

495 Shelton, 1998; Carlson, et al., 2013; Cichy, et al., 2014; Kiani, et al., 2007; Kriegeskorte, Mur,  
496 Ruff, et al., 2008; Proklova, et al., 2016). Our study included stimuli that do not clearly have  
497 membership in the animate or inanimate categories (Bracci & Op de Beeck, 2016; Carlson, et al.,  
498 2013; Cichy, et al., 2014; Konkle & Caramazza, 2013; Kriegeskorte, Mur, Ruff, et al., 2008;  
499 Proklova, et al., 2016). The poor performance of the animate vs. inanimate model (and similarly the  
500 living vs. non-living model) likely could be accounted for by the inclusion of robots and toys. For  
501 example, visually inspecting the t-SNE plots (Figure 4) shows that robots are represented closer to  
502 humans and animate objects than to inanimate objects. This suggests that an animate/inanimate  
503 distinction is not the best way to classify these stimuli, and further highlights the impact of stimulus  
504 selection on defining the organisation of object categories (c.f. Carlson, Goddard, Kaplan, Klein, &  
505 Ritchie, 2018; Goddard, Klein, Solomon, Hogendoorn, & Carlson, 2018). Indeed, a recent fMRI  
506 study by Bracci, Kalfas, & Op de Beeck (2017) showed that visually confusing objects (e.g., a mug  
507 in the shape of a cow) exhibited neural activity patterns that were more similar to animate objects  
508 (i.e., an actual cow) than inanimates (Bracci, Kalfas, & Op de Beeck, 2017). Furthermore, as  
509 exemplar typicality affects the distinctiveness of category representations (Iordan, Greene, Beck, &  
510 Fei-Fei, 2016), the inclusion of these ambiguous object categories may have disproportionately  
511 affected a strict dichotomous categorisation model.

512 The strong performance of the human model in late stage processing builds on our existing  
513 understanding of the representation of object categories in the brain. [This aligns with recent work](#)  
514 [demonstrating the role of higher order constructs like agency in the visual representation of objects](#)  
515 ([Haxby, et al., 2020](#)); and extends on the continuum idea of Connolly et al. (2012) and Sha et al.  
516 ([2015](#)), as it represents a type of human-similarity continuum (see also Thorat, et al., 2019). Unlike  
517 the animacy continuum that is based on biological classes, the human model was not limited by  
518 biology (Gobbini, et al., 2011; Tong, et al., 2000). Results from an fMRI study by Gobbini et al.  
519 (2011) are also consistent with a level of cross-over between animate/inanimate object categories  
520 that does not fit into this dichotomy, nor a continuum based on biological classes. The authors

521 compared human observers' perception of human faces and robots and found that robots evoked  
522 activation in areas associated with faces (though to a lesser extent than humans), while also  
523 activating object areas and areas associated with mechanical movements. This supports the idea of  
524 more a complex model of object categorisation that incorporates factors such as agency and human-  
525 related experiences. Given the relative strength of our human-centric model in accounting for the  
526 data, the idea of "humanness" as an important dimension in the neural representation of objects  
527 warrants further exploration.

528 Our study tested a diverse range of stimuli and models. The richness of the dataset offers the  
529 future possibility of testing additional models. For example, one could envision a hybrid face model  
530 that indicates the presence of a face in the image with levels based on species: human, non-human  
531 primate, mammals, insects, inanimate objects with faces (toys, robots). The dataset accordingly will  
532 be made publicly available for future work. The diverse set of models tested has both limitations  
533 and benefits. One limitation is many of the models are not independent (Figure 2E), which creates  
534 issues in interpretability. For example, the fits of the face/body model and the shape model to the  
535 MEG data had similar trajectories (Figure 3B and 3C). The two models are also correlated (2E), so  
536 it remains unknown whether the explanatory power of each of the models is based on the presence  
537 of face or body in the image or the shape of the figure in the image. Issues such as these will need to  
538 be disentangled in future work using a stimulus set designed to target these individual research  
539 questions. One strength of testing such a wide range of models is each individual models' fit to the  
540 data gives insight into the brain's representations of the stimuli. If, for example, we had tested the  
541 human similarity model in the absence of the agency and experience models, the nature of the  
542 humanness model would be considerably more abstract. Here, we can see from the overlap between  
543 the agency/experience models (Figure 2E) that the notion of humanness is captured well by factors  
544 like capacity for agency and experience.

545 Our human-centric model likely encompasses a complex set of features, including both  
546 visual and conceptual factors. This is evidenced by the overlap between the human-centric model

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

558 Presently, we still do not have a clear understanding of how different semantic concepts  
559 relate to object representations and category structure. A recent model attempts to explain the neural  
560 representation of object attempts using a multidimensional framework (Martin, 2016). In this paper,  
561 the author suggests that neural patterns associated with objects are formed from complex interactive  
562 circuits based on a range of systems throughout the brain, including those associated with action,  
563 perception and emotion. This idea shifts the focus away from models based on categories, with a  
564 view to a more holistic approach to object representations that considers interactions between  
565 various circuits throughout the brain. In this multidimensional framework, it is essential to  
566 recognize that no single feature or attribute could be able to fully explain the richness of the brain’s  
567 (multidimensional) representation of objects (see e.g., Thorat, et al., 2019). Recent fMRI studies  
568 have sought to identify principle axes of object representations in the brain (e.g., Connolly, et al.,  
569 2012; Sha, et al., 2015; Thorat, et al., 2019). In the present study, we show that the human  
570 similarity provided a strong account of late stage processing, highlighting “humanness” as a key  
571 feature in the human brain’s representation of objects that shapes our experience of the world.

572

573 **Acknowledgements:**  
574 This research was supported by an Australian Research Council Future Fellowship (FT120100816)  
575 and an Australian Research Council Discovery project (DP160101300) awarded to T.A.C. The  
576 authors acknowledge the University of Sydney HPC service for providing High Performance  
577 Computing resources. The authors declare no competing financial interests.  
578

579 **Author contributions:**  
580 E.C, E.G, M.W, T.C designed the study  
581 E.C, T.G collected the data  
582 E.C, E.G, T.G conducted the analysis  
583 E.C, E.G, T.G, T.C interpreted the results  
584 E.C and T.C wrote the manuscript  
585 All authors reviewed the manuscript  
586

587 **References**

- 588 Barragan-Jason, G., Cauchoux, M., & Barbeau, E. J. (2015). The neural speed of familiar face  
589 recognition. *Neuropsychologia*, 75, 390-401.
- 590 Bracci, S., Kalfas, I., & Op de Beeck, H. (2017). The ventral visual pathway represents animal  
591 appearance over animacy, unlike human behavior and deep neural networks. *BioRxiv*,  
592 228932.
- 593 Bracci, S., & Op de Beeck, H. (2016). Dissociations and Associations between Shape and Category  
594 Representations in the Two Visual Pathways. *J Neurosci*, 36, 432-444.
- 595 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, 10, 433-436.
- 596 Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the  
597 animate-inanimate distinction. *J Cogn Neurosci*, 10, 1-34.
- 598 Carlson, T. A., Goddard, E., Kaplan, D. M., Klein, C., & Ritchie, J. B. (2018). Ghosts in machine  
599 learning for cognitive neuroscience: Moving from data to theory. *Neuroimage*, 180, 88-100.
- 600 Carlson, T. A., Simmons, R. A., Kriegeskorte, N., & Slevc, L. R. (2014). The emergence of semantic  
601 meaning in the ventral temporal pathway. *J Cogn Neurosci*, 26, 120-131.
- 602 Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of  
603 object vision: the first 1000 ms. *J Vis*, 13.
- 604 Cauchoux, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The neural dynamics of face  
605 detection in the wild revealed by MVPA. *J Neurosci*, 34, 846-854.
- 606 Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex  
607 for perceiving and knowing about objects. *Nat Neurosci*, 2, 913-919.
- 608 Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal  
609 stream. *Neuroimage*, 12, 478-484.
- 610 Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural  
611 networks to spatio-temporal cortical dynamics of human visual object recognition reveals  
612 hierarchical correspondence. *Sci Rep*, 6, 27755.
- 613 Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and  
614 time. *Nat Neurosci*, 17, 455-462.
- 615 Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *J  
616 Neurosci*, 34, 4766-4775.
- 617 Coggan, D. D., Baker, D. H., & Andrews, T. J. (2016). The Role of Visual and Semantic Properties in  
618 the Emergence of Category-Specific Patterns of Neural Response in the Human Brain.  
619 *eNeuro*, 3.
- 620 Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., Abdi, H., & Haxby,  
621 J. V. (2012). The representation of biological classes in the human brain. *J Neurosci*, 32,  
622 2608-2618.
- 623 Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object  
624 recognition in the human brain: From visual features to categorical decisions.  
625 *Neuropsychologia*, 105, 165-176.
- 626 DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?  
627 *Neuron*, 73, 415-434.
- 628 Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain  
629 specificity in visual cortex. *Cereb Cortex*, 16, 1453-1461.
- 630 Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual  
631 processing of the human body. *Science*, 293, 2470-2473.
- 632 Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area:  
633 recognition, navigation, or encoding? *Neuron*, 23, 115-125.

- 634 Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment.  
635 *Nature*, 392, 598-601.
- 636 Gainotti, G. (2000). What the locus of brain lesion tells us about the nature of the cognitive defect  
637 underlying category-specific disorders: a review. *Cortex*, 36, 539-559.
- 638 Gobbini, M. I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., Guazzelli, M., & Pietrini, P.  
639 (2011). Distinct neural systems involved in agency and animacy detection. *J Cogn Neurosci*,  
640 23, 1911-1920.
- 641 Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016). Representational dynamics of  
642 object recognition: Feedforward and feedback information flows. *Neuroimage*, 128, 385-  
643 397.
- 644 Goddard, E., Klein, C., Solomon, S. G., Hogendoorn, H., & Carlson, T. A. (2018). Interpreting the  
645 dimensions of neural feature representations revealed by dimensionality reduction.  
646 *Neuroimage*, 180, 41-67.
- 647 Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315, 619.
- 648 Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from  
649 Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series  
650 Neuroimaging Data. *J Cogn Neurosci*, 29, 677-697.
- 651 Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed  
652 and overlapping representations of faces and objects in ventral temporal cortex. *Science*,  
653 293, 2425-2430.
- 654 Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for  
655 agentic action in visual representation. *Neuroimage*, 116561.
- 656 Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The  
657 functional organization of human extrastriate cortex: a PET-rCBF study of selective  
658 attention to faces and locations. *J Neurosci*, 14, 6336-6353.
- 659 Haynes, J. D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and  
660 Perspectives. *Neuron*, 87, 257-270.
- 661 Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes  
662 the representation of thousands of object and action categories across the human brain.  
663 *Neuron*, 76, 1210-1224.
- 664 Jordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2016). Typicality sharpens category  
665 representations in object-selective cortex. *Neuroimage*, 134, 170-179.
- 666 Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed  
667 representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A*,  
668 96, 9379-9384.
- 669 Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des  
670 Jura. *Bulletin de la Société vaudoise des sciences naturelles*, 37.
- 671 Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human  
672 extrastriate cortex specialized for face perception. *J Neurosci*, 17, 4302-4311.
- 673 Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response  
674 patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol*, 97,  
675 4296-4309.
- 676 Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36, 14-14.
- 677 Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and  
678 object size. *J Neurosci*, 33, 10235-10242.
- 679 Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition,  
680 computation, and the brain. *Trends Cogn Sci*, 17, 401-412.
- 681 Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting  
682 the branches of systems neuroscience. *Front Syst Neurosci*, 2, 4.

- 683 Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P.  
684 A. (2008). Matching categorical object representations in inferior temporal cortex of man  
685 and monkey. *Neuron*, *60*, 1126-1141.
- 686 Maaten, L. v. d., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning  
687 Research*, *9*, 2579–2605.
- 688 Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumia, R. I., Caramazza, A., & Martin, A. (2007). Action-  
689 related properties shape object representations in the ventral stream. *Neuron*, *55*, 507-  
690 520.
- 691 Martin, A. (2016). GRAPES-Grounding representations in action, perception, and emotion systems:  
692 How object properties and categories are represented in the human brain. *Psychon Bull  
693 Rev*, *23*, 979-990.
- 694 Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox  
695 for representational similarity analysis. *PLoS Comput Biol*, *10*, e1003553.
- 696 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into  
697 movies. *Spat Vis*, *10*, 437-442.
- 698 Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial  
699 overview. *Neuroimage*, *45*, S199-209.
- 700 Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S. (2009).  
701 Neural representations of faces and body parts in macaque and human cortex: a  
702 comparative fMRI study. *J Neurophysiol*, *101*, 2581-2600.
- 703 Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and  
704 Object Category in Human Visual Cortex: The Animate-Inanimate Distinction. *J Cogn  
705 Neurosci*, *28*, 680-692.
- 706 Proklova, D., Kaiser, D., & Peelen, M. V. (2019). MEG sensor patterns reflect perceptual but not  
707 categorical similarity of animate and inanimate objects. *Neuroimage*, *193*, 167-177.
- 708 Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human  
709 visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging  
710 study. *J Neurosci*, *16*, 5205-5215.
- 711 Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat  
712 Neurosci*, *2*, 1019-1025.
- 713 Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object  
714 processing. A positron emission tomography study. *Brain*, *115 Pt 1*, 15-36.
- 715 Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory  
716 of immediate visual recognition. *Prog Brain Res*, *165*, 33-56.
- 717 Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C.  
718 (2015). The animacy continuum in the human ventral vision pathway. *J Cogn Neurosci*, *27*,  
719 665-678.
- 720 Taylor, J. C., & Downing, P. E. (2011). Division of labor between lateral and ventral extrastriate  
721 representations of faces, bodies, and objects. *J Cogn Neurosci*, *23*, 4122-4137.
- 722 Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human  
723 ventral temporal cortex. *arxiv.org, arXiv:1904.02866*.
- 724 Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*,  
725 *381*, 520-522.
- 726 Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties  
727 of the human fusiform face area. *Cogn Neuropsychol*, *17*, 257-280.
- 728 van de Nieuwenhuijzen, M. E., Backus, A. R., Bahramisharif, A., Doeller, C. F., Jensen, O., & van  
729 Gerven, M. A. (2013). MEG-based decoding of the spatiotemporal dynamics of visual  
730 category perception. *Neuroimage*, *83*, 1063-1073.

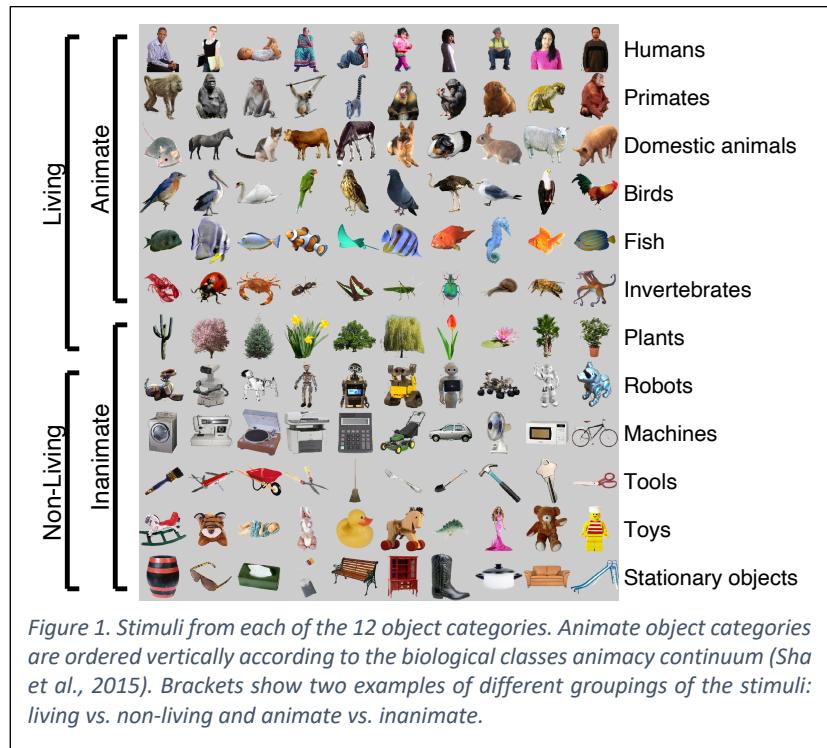
731 Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107 ( Pt 3),  
732 829-854.  
733

734

735

736

737



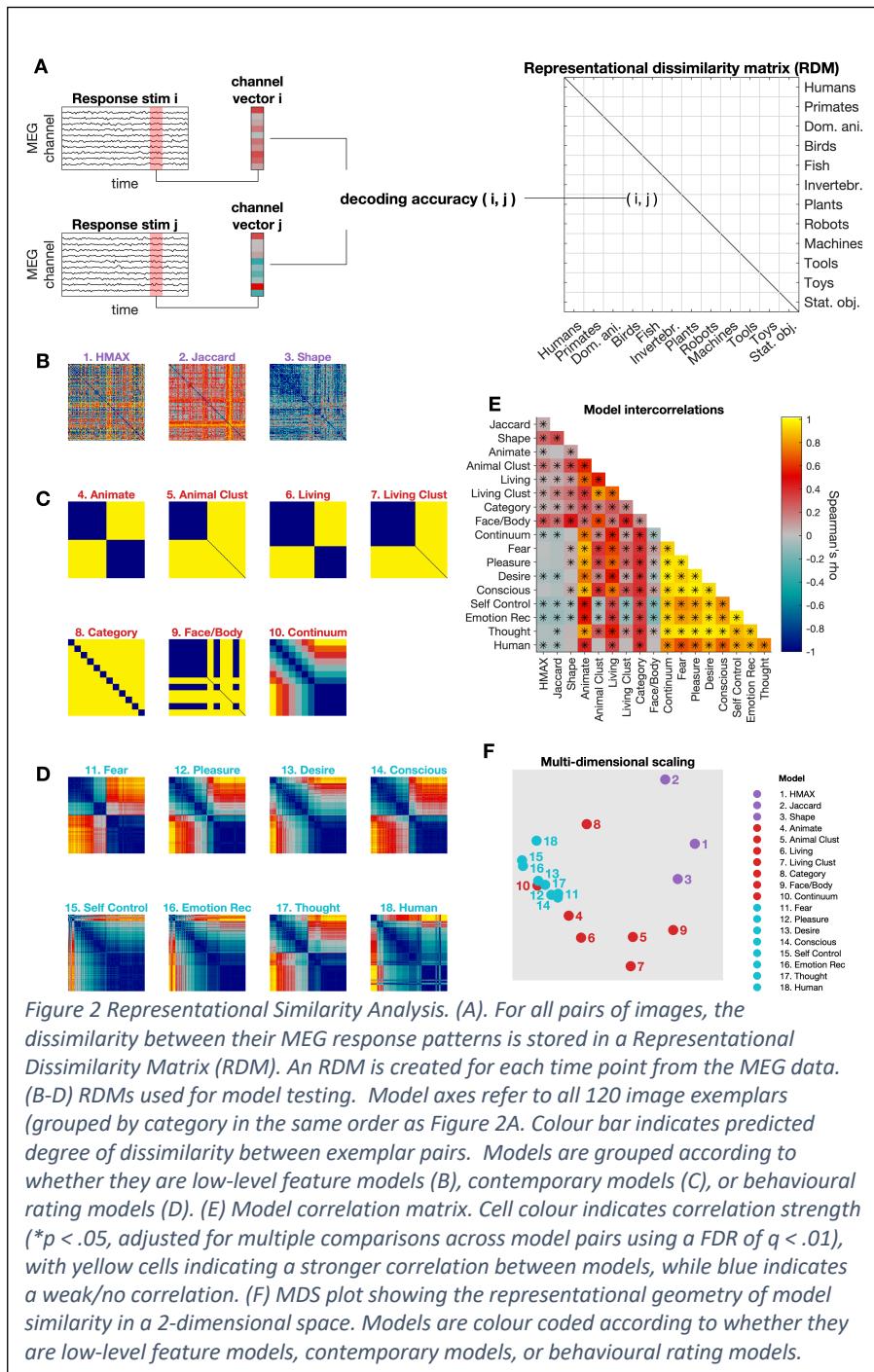
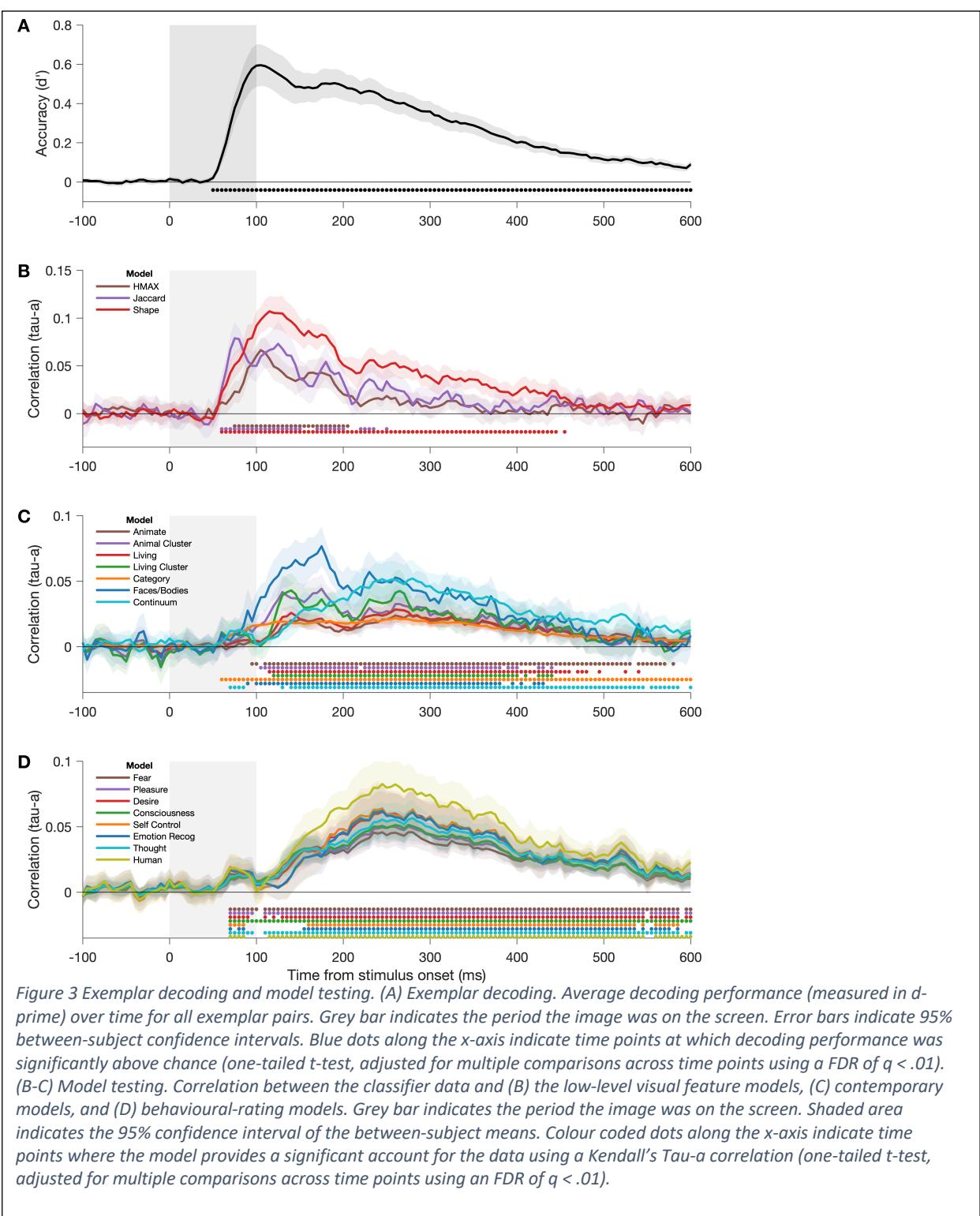
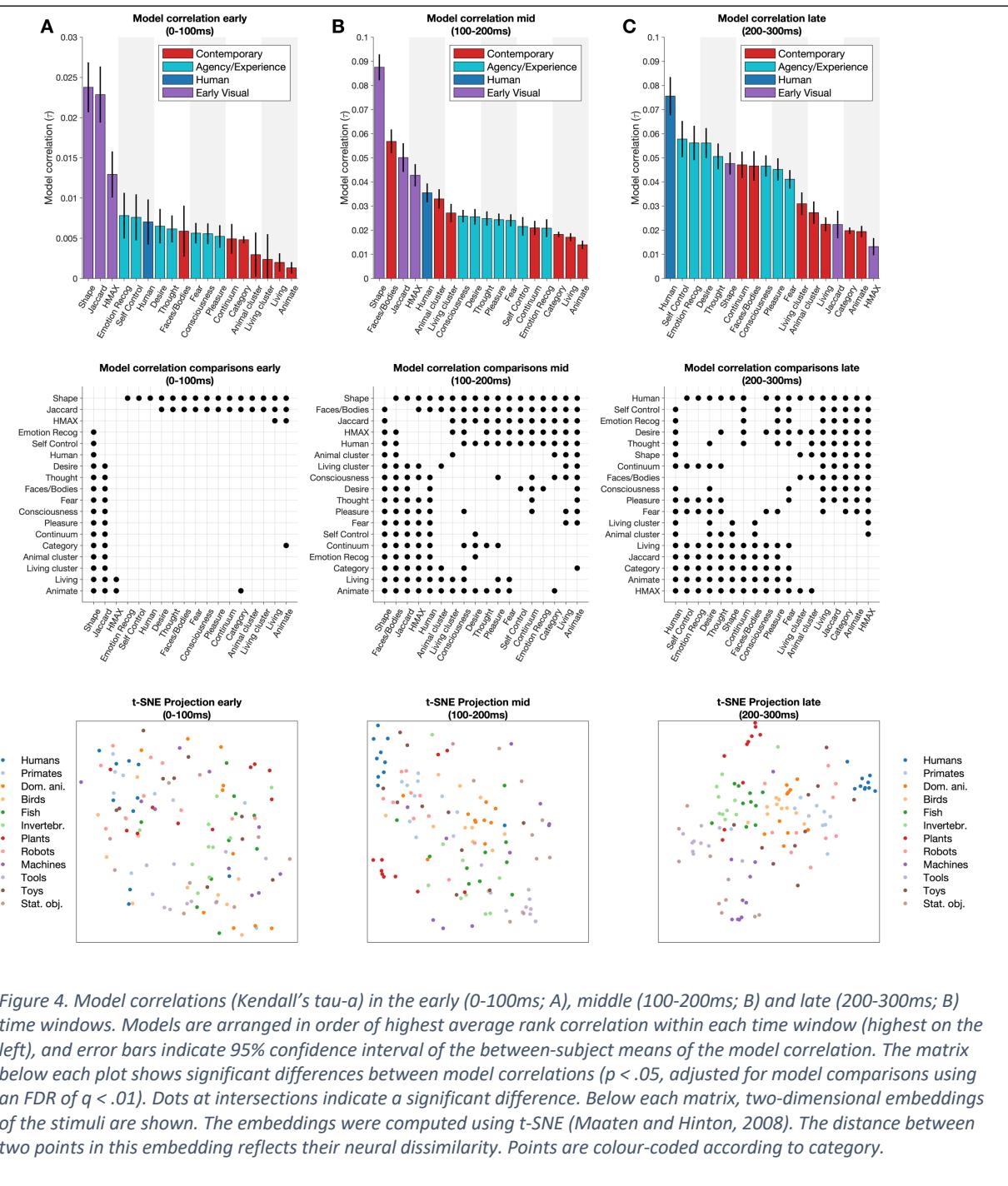


Figure 2 Representational Similarity Analysis. (A) For all pairs of images, the dissimilarity between their MEG response patterns is stored in a Representational Dissimilarity Matrix (RDM). An RDM is created for each time point from the MEG data. (B-D) RDMs used for model testing. Model axes refer to all 120 image exemplars (grouped by category in the same order as Figure 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  $q < .01$ ), 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.





Erika Contini: Conceptualization, Formal analysis, Roles/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, Roles/Writing - original draft, Writing - review & editing.

Data will be made publicly available after acceptance.