

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. McGill Vision Research, Department of Ophthalmology, McGill University, Montreal, Canada

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

9

10   **Corresponding author:**

11   Thomas Carlson

12   School of Psychology

13   University of Sydney, 2050 NSW Australia

14

15

16

17    **Abstract**

18    Neuroimaging studies investigating human object recognition have largely 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 hypothesised dichotomies, for example animate  
21    versus inanimate, and continuous feature dimensions, such as biologically similarity. These studies  
22    typically have used stimuli that are clearly identified as animate or inanimate, neglecting objects  
23    that may not fit into this dichotomy. We generated a novel stimulus set including standard objects  
24    and objects that blur the animate-inanimate dichotomy, for example robots and toy animals. We  
25    used 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 brain  
29    responses are best accounted for by low-level visual similarity of the objects; and shortly thereafter,  
30    higher order models of agency/experience best explained the brain's representation of the stimuli.  
31    Strikingly, a model of human-similarity provided the best account for the brain's representation  
32    after an initial perceptual processing phase. Our findings provide evidence for a new dimension of  
33    object coding in the human brain – one that has a “human-centric” focus.

34 **Introduction**

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

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

60 representation space. Animate subcategories have been argued to be coded along an axis of  
61 biologically similarity to humans (Connolly, et al., 2012; Sha, et al., 2015). This animacy  
62 continuum, however, does not provide a clear prediction for subcategory differentiation within the  
63 inanimate domain, nor for how the brain would represent objects that blur the animate-inanimate  
64 distinction (e.g., robots and animal toys). Moreover, it is also unclear whether a continuum centred  
65 around ‘animacy’ best captures the dimension along which neural responses vary. Sha et al. (2015),  
66 for example, proposed that the neural representation of objects is better characterised according to  
67 the object’s ability to perform goal-directed actions (see also Thorat, Proklova, & Peelen, 2019).  
68 Critically, there are many related factors to biologically similarity and agency that are known to  
69 influence human perception of objects (Gobbini, et al., 2011; Gray, Gray, & Wegner, 2007). This  
70 raise the question about whether these factors also might be used as organisational principles for the  
71 brain’s representation of objects.

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

81

82 **Materials and Methods**

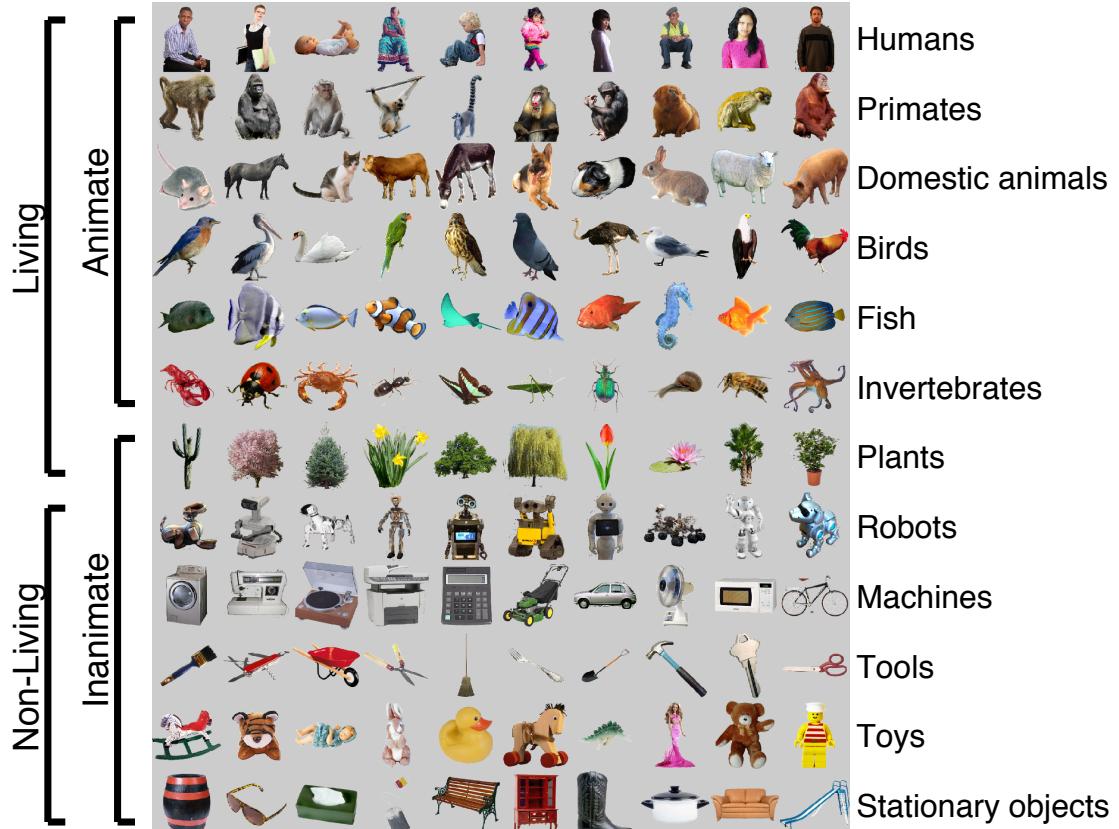
83 ***Participants***

84 Twenty-four English-speaking volunteers (18 female) with an average age of 24.7 years (SD =  
85 5.47; range = 18-37) were recruited from the Macquarie University community. Informed written  
86 consent was obtained prior to participation, and participants were financially compensated for their  
87 time. All participants self-reported normal or corrected-to-normal vision (wearing of contacts was  
88 allowed), were free of medical conditions, and were not currently taking any neuroactive  
89 medications. This study was approved by the Macquarie University Human Research Ethics  
90 Committee.

91

92 ***Stimuli***

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



104

105 **Figure 1.** Stimuli from each of the 12 object categories. Animate object categories are ordered  
106 vertically according to the biological classes animacy continuum (Sha et al., 2015). Brackets show  
107 two examples of different groupings of the stimuli: living vs. non-living and animate vs. inanimate.  
108

109 **MEG Experimental Procedure**

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

117 Across trials, object images were manipulated in two ways to reduce the effects of low-level  
118 stimulus properties on our data. Firstly, a left-right flipped version of each image was included in  
119 the stimulus set, resulting in a total of 240 stimuli from 120 object images. Secondly, during image

120 presentation, stimuli appeared in one of four locations while participants maintained fixation on a  
121 central marker, thus varying retinal location of the stimulus images. The four locations were defined  
122 by a shift from central presentation towards each of the four corners of the screen, where each  
123 stimulus location overlapped the central fixation point (details in *Display Apparatus* below). Each  
124 stimulus was presented three times at each location. This resulted in a total of 2880 trials (240  
125 stimuli x 4 locations x 3 repetitions = 2880 trials). The additional trials were not included in the  
126 analysis: these included the first and last trial of each block, as well as 288 repeat trials that were  
127 added for the attention task (see below).

128

129 ***Attention Task***

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

138

139 ***Display Apparatus***

140 Participants lay supine in the magnetically shielded recording room. Using an InFocus IN5108  
141 projector situated outside the chamber, stimuli were projected onto a mirror, which reflected the  
142 image onto the ceiling, located approximately 113 cm above the participant. The total screen area  
143 was 20x15 degrees of visual angle (DVA). Throughout the experiment the screen background was  
144 held at a mean grey, and subjects were instructed to fixate on a black central fixation point  
145 (diameter of 0.1 DVA) that was always present. All stimulus locations were within a 6.9 DVA

146 square, centred on the fixation point. Each stimulus consisted of a 256x256 pixel image (containing  
147 the segmented colour object) that was drawn to a 4.9x4.9 DVA square. Stimuli were presented one  
148 at a time, in one of four locations aligned with the upper left, upper right, lower left, or lower right  
149 corner of the 6.9 DVA square. A central square of 150 pixels (2.9 DVA) was common to all four  
150 stimulus locations. All stimuli were drawn as full colour segmented objects against a mean grey  
151 background (as in Figure 1): the same mean grey as the screen outside the stimulus location. Upon  
152 stimulus presentation, a 50x50 pixel (1x1 DVA) white square simultaneously appeared in the  
153 bottom right corner of the projection, which was aligned with a photodetector attached to the mirror  
154 to accurately record the stimulus presentation time in the MEG recording. The experiment was run  
155 on a Dell PC desktop computer using MATLAB software (Natick, MA) and the Psychophysics  
156 Toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

157

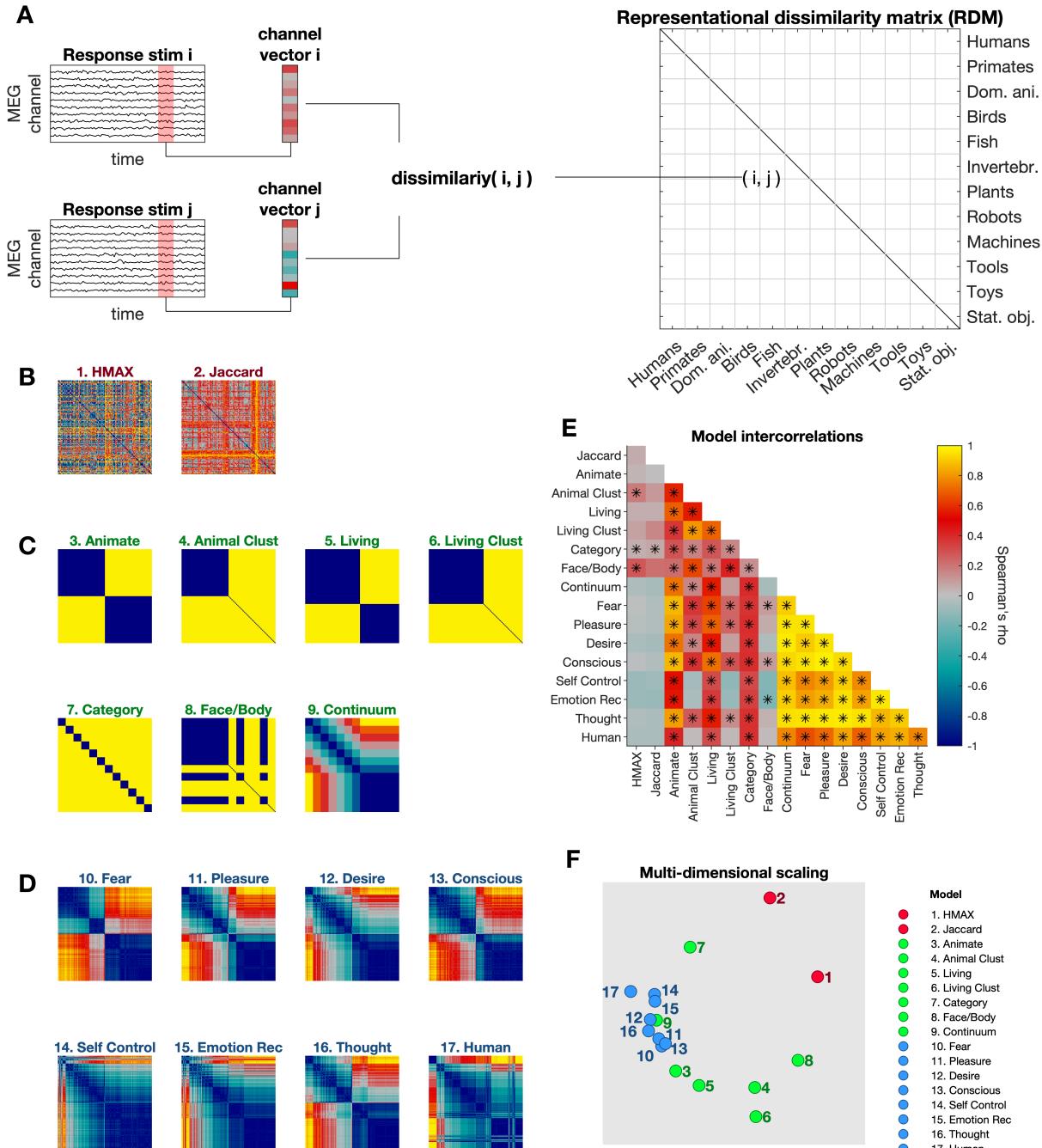
158 **MEG Data Acquisition** MEG data were recorded in the KIT-Macquarie Brain Research Laboratory  
159 using a 160-channel whole-head axial gradiometer (KIT, Kanazawa, Japan). Continuous data were  
160 acquired at a sampling rate of 1000 Hz, and were band-pass-filtered online from 0.03 to 200 Hz.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

167

168 **Classification analysis** For each participant, we used linear discriminant analysis to classify  
169 object/exemplar identity at the single trial level, training and testing classifiers on their ability to  
170 discriminate every possible exemplar pair of the 120 object images. We used cross-validated  
171 classification accuracy as a measure of how dissimilar the patterns of brain activity were for one

172 exemplar compared to another (Nili, et al., 2014). We did not attempt to model the effects of spatial  
173 position or left-right flip in our classification analysis, but instead used a single data label (the  
174 object identity) for data obtained from both the standard and left-right-flipped versions of the  
175 stimuli, as well as all four stimulus presentation locations. By including data from all variations of  
176 the stimuli, we sought to force the classifier to generalise beyond lower-level visual features, (such  
177 as the presence or absence of stimulation at a given location in the visual field), and instead use any  
178 neural correlate of object identity. These modifications to the stimulus presentation would have  
179 introduced extra noise into the signal across trials, so would tend to reduce classifier performance  
180 relative to unvarying stimuli, but they allowed us to better target higher-level object representations.  
181 For each time-point, we trained and tested a separate classifier to discriminate each pair of exemplar  
182 identities from the PCA components. We used a 10-fold cross-validation procedure, where the  
183 classifier was trained on data from 90% of the trials and then its accuracy was evaluated using its  
184 performance when classifying the remaining 10% of the data, so that the classifier was never tested

185 on data that were included in the training set. This process was repeated 10 times, so that all trials  
 186 were used as test data once each. D-prime ( $d'$ ) was used as the metric for classification accuracy.



187

188 **Figure 2 Models. Representational Similarity Analysis (A).** For all pairs of images, the dissimilarity  
 189 between their MEG response patterns is stored in a Representational Dissimilarity Matrix (RDM).  
 190 An RDM is created for each time point from the MEG data. *RDMs used for model testing (B-D).*  
 191 Model axes refer to all 120 image exemplars (grouped by category in the same order as Figure 2A).  
 192 Colour bar indicates predicted degree of dissimilarity between exemplar pairs. Models are grouped  
 193 according to whether they are low-level feature models (B), contemporary models (C), or  
 194 behavioural models (D). *Model intercorrelations (E-F).* E) Model correlation matrix. Cell colour  
 195 indicates correlation strength ( $*p < .05$ , adjusted for multiple comparisons across time points using  
 196 a FDR of  $q < .01$ ), with yellow cells indicating a stronger correlation between models, while blue  
 197 indicates a weak/no correlation. F) MDS plot showing the representational geometry of model

198 similarity in a 2 dimensional space. Models are colour coded according to whether they are low-  
199 level feature models, contemporary models, or behavioural models.  
200 **Representational Similarity Analysis (RSA)** Classifier accuracies ( $d'$ ) were averaged across  
201 exemplar pairs to obtain the mean classifier performance for each time point. Additionally, to  
202 capture the pattern of classifier performance across exemplar pairs and compare this pattern with  
203 model predictions, we constructed a Representational Dissimilarity Matrix (RDM) for each time  
204 point. The RDM is a 120x120 matrix, symmetric along the diagonal, where each cell is the  
205 classification accuracy ( $d'$ ) for that pair of exemplars.

206 For each time point we compared each participant's observed neural RDM with model  
207 RDMs, where each model RDM was a 120x120 matrix derived from theory, computational  
208 modelling, or behavioural data (as described in detail below). This analysis, known as  
209 'Representational Similarity Analysis' (RSA) (Kriegeskorte, Mur, & Bandettini, 2008) tests the  
210 relationship between models of interest and the group data, measuring how well the model RDMs  
211 account for the observed pattern of results. At each time point we used Kendall's tau-a to compute  
212 the rank order correlation between each candidate model and the neural data, and used these  
213 correlation values to compare candidate models in their ability to account for the neural data. Figure  
214 2 shows the model RDMs, which are described in detail below.

215

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

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

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

224 layer (C2) for every stimulus were vectorized. We then generated the model RDM by taking the  
225 Euclidean distance between the vectorized model responses for each pair of stimuli.  
226 *Jaccard (silhouette model; model 2)*: An abstract shape model that measures the shape of each  
227 object in terms of the pixels that the image occupies (Jaccard, 1901). We generated the model RDM  
228 by comparing the overlapping silhouette regions of two images at a time and obtaining a measure of  
229 the difference. This model was generated based on the standard orientation of each stimulus (i.e.,  
230 not flipped), independent of location.

231

232 ***Contemporary models of object representations (Figure 2, models 3-9)***

233 The contemporary models were created based on organisational structures proposed in previous  
234 studies, with the term ‘contemporary’ used to highlight that these reflect current theories of object  
235 category structure. Descriptions of each model are provided below. *Dichotomy models (models 3*  
236 *and 5)*: The animate vs. inanimate dichotomy model (Caramazza & Shelton, 1998; Carlson, et al.,  
237 2013; Cichy, Pantazis, & Oliva, 2014; Kriegeskorte, Mur, Ruff, et al., 2008) is a category model  
238 that grouped all animate and inanimate objects separately (implying that objects within these  
239 groupings were more similar to each other, and more dissimilar to objects in the other grouping).  
240 Similarly, the living vs. non-living dichotomy model (Gainotti, 2000; Huth, Nishimoto, Vu, &  
241 Gallant, 2012; Warrington & Shallice, 1984) grouped all living and non-living objects separately.  
242 The living category included the same items as the animate category but with the addition of plants.  
243 *Cluster models (4 and 6)*: The animal cluster model (model 4) is a single-category model that only  
244 grouped all animate objects together, suggesting that animate objects will be more similar to each  
245 other, and more dissimilar to all other objects, but that inanimate objects will not cluster. The living  
246 cluster model (model 6) follows the same principle, but grouping all living objects together. The  
247 cluster models were created to determine whether the effect of the dichotomy models was driven by  
248 cohesion within the in-group alone (i.e., animate, living), with more disparate object representations  
249 in the out-group category (i.e., inanimate, non-living) (Clarke & Tyler, 2014).

250     *Faces/bodies model (model 7):* Faces and bodies stand out as special categories for object  
251     recognition (Barragan-Jason, Cauchoix, & Barbeau, 2015; Cauchoix, Barragan-Jason, Serre, &  
252     Barbeau, 2014; Gobbini, et al., 2011; Haxby, et al., 2001; van de Nieuwenhuijzen, et al., 2013) and  
253     so were of interest given the inclusion of toys and robots in our stimulus set. As such, the  
254     faces/bodies model is single-category model, grouping together all object categories that had faces  
255     or bodies, including all animate objects, as well as robots and toys.

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

261     *Continuum model (model 9):* The continuum model is a graded model based on the animacy  
262     continuum proposed by Sha et al. (2015). The continuum included a gradient of similarity between  
263     object categories that varied along a dimension related to biological classes, such that categories  
264     more similar to humans (biologically), would have more similar activity patterns, and those more  
265     dissimilar to humans would have activity patterns more similar to inanimate objects. For this model,  
266     plants were included on the continuum as they are a biological category and were represented on  
267     the continuum between invertebrates and inanimate objects. All non-living inanimate objects were  
268     treated as a single category, most dissimilar to the human category.

269

270     *Behavioural-rating models (Figure 2, models 10-17)*

271     The behavioural-rating models include the agency/experience models (models 10 – 16) and the  
272     human model (model 17). These models were created by obtaining behavioural ratings of the  
273     stimuli according to a specific question (detailed below). A total of 325 Amazon's Mechanical Turk  
274     workers residing in either the United States of America or Canada, completed one of the eight  
275     surveys online (number of participants per survey ranged from 40 – 43). Participants included 146

276 females (1 other, 1 no response), and had an average age of 35.27 years (SD = 10.26, range = 18.9 –  
277 70.8; one age value missing). In each survey we asked workers to answer a single question for each  
278 of the stimuli:

279

280 10. *Fear* – How much is it capable of feeling afraid or fearful?

281 11. *Pleasure* – How much is it capable of experiencing physical or emotional pleasure?

282 12. *Desire* - How much is it capable of longing or hoping for things?

283 13. *Consciousness* - How much is it capable of having experiences and being aware of things?

284 14. *Thought* - How much is it capable of thinking?

285 15. *Emotion-recognition* - How much is it capable of understanding how others are feeling?

286 16. *Self-Control* - How much is it capable of exercising self-restraint over desires, emotions or  
287 impulses?

288 17. *Human* – How similar is this to a human?

289

290 Surveys for models 10-16 were based on a subset of the mental capacity surveys used in Gray et al.  
291 (2007), which vary as to how much they loaded onto the author's 'Experience' and 'Agency'  
292 factors that were established in their study. The seven agency/experience models were based on the  
293 results of these surveys. The 'Human' survey (17) was added to address a meta-representational  
294 idea of categorization, that of "human-ness": a complex factor which may encompass biology,  
295 agency, and visual similarity. Each survey required participants to rate all 120 images on a 7-point  
296 scale from 'Not at all' to 'Very much so' in response to the specific question. Each survey took  
297 approximately 10 minutes to complete and participants were financially compensated for their time.  
298 The surveys were created and administered using the Qualtrics online survey platform. For each  
299 survey, participants provided voluntary consent and basic demographic information before  
300 completing the survey. Participants were only allowed to complete one of the eight surveys

301 available, resulting in unique individuals for each survey. Stimulus order was randomised  
302 separately for each participant.  
303 To construct the models based on agency and experience (shown in Figure 2), an RDM was created  
304 for each set of survey responses by obtaining the absolute difference between image ratings for each  
305 pairwise comparison of the 120 images, using the mean ratings of each image. These RDMs, based  
306 on the survey ratings, provide hypothetical models of the degree of dissimilarity between the neural  
307 responses associated with each image. For graphical purposes, we scaled these difference values  
308 between 0 and 1 for each model, such that warmer colours indicate greater dissimilarity, while  
309 cooler colours depict greater similarity between the neural representations in the pair-wise  
310 comparison.

311

312 ***Model intercorrelations (Figure 2E)***

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

320 In this study, we aimed to select stimuli that were visually diverse within each subcategory, to  
321 minimise the extent to which visual similarity would produce seemingly ‘categorical’ patterns of  
322 results. The model correlation data suggests that our stimulus set provided good separation of visual  
323 similarity and object category, since few models correlated with the visual feature models.

324 Importantly, this should minimise the contribution of low-level visual similarity when we evaluate  
325 our hypothesis driven models. Exceptions to this included the animal cluster, category, and  
326 faces/bodies models, which each showed a significant correlation with one, or both of the HMAX

327 and Jaccard models. This suggests that despite our stimulus diversity, there was still greater visual  
328 homogeneity of exemplars within the category groupings in these models than between category  
329 groupings. This means that, particularly for the animal cluster, category, and faces/bodies models,  
330 any correlation between these models and the observed pattern of classifier performance could be  
331 driven by low-level visual similarity rather than by the higher-level category structure represented  
332 by these models.

333

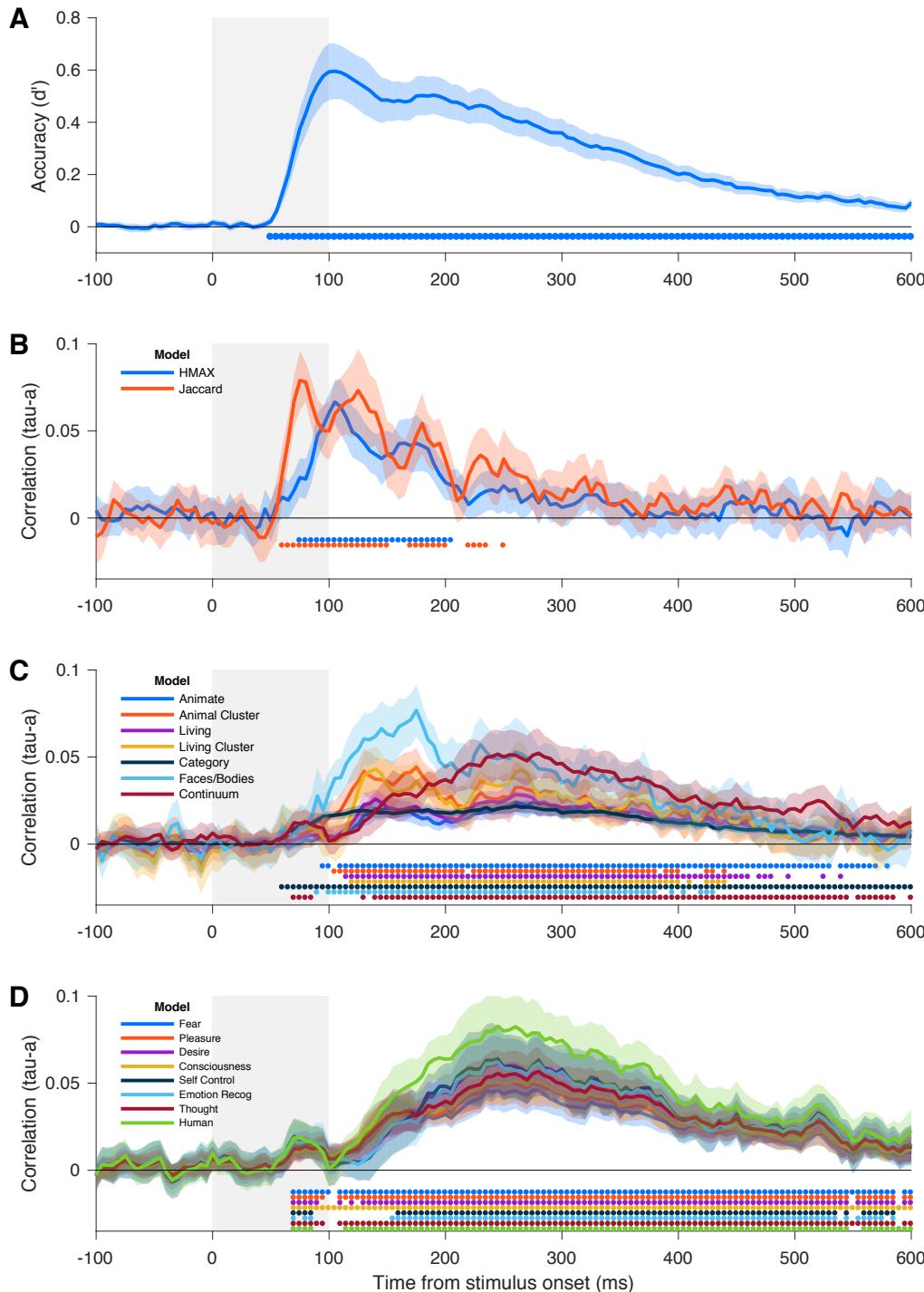
334 **Results**

335

336 **Decoding object exemplars from the MEG recordings**

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

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



350

351 **Figure 3. Exemplar decoding and model testing. (A) Exemplar decoding.** Average decoding  
 352 performance (measured in  $d'$ ) over time for all exemplar pairs. Grey bar indicates the period  
 353 the image was on the screen. Error bars indicate 95% between-subject confidence intervals. Blue  
 354 dots along the x-axis indicate time points at which decoding performance was significantly above  
 355 chance (one-tailed t-test, adjusted for multiple comparisons across time points using a FDR of  $q <$   
 356 .01). **(B-C) Model testing.** Correlation between the classifier data and (B) the low-level visual  
 357 feature models, (C) contemporary models, and (D) behavioural-rating models. Grey bar indicates  
 358 the period the image was on the screen. Shaded area indicates the 95% confidence interval of the  
 359 between-subject means. Colour coded dots along the x-axis indicate time points where the model  
 360 provides a significant account for the data using a Kendall's Tau-a correlation (one-tailed t-test,  
 361 adjusted for multiple comparisons across time points using a FDR of  $q < .01$ ).

362 **The dynamic representation of objects**

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

371

372 **Low-level visual models account for early representations (Figure 3B).**

373 We evaluated two low-level models to study how primitive visual features account for the brain's  
374 representation of the stimuli. The Jaccard (i.e., silhouette) model evaluates the global shape of the  
375 stimuli (Jaccard, 1901). The HMAX model is based on a simulation of the response of early visual  
376 areas (Serre et al., 2007). Both the Jaccard and HMAX models were significantly correlated with  
377 the neural RDMs during early stages in the time course, peaking at 75 and 105 ms respectively, and  
378 were no longer significant predictors after 250 ms. This is in agreement with the established  
379 literature about the time-course of visual object recognition, with responses related to lower-level  
380 visual stimulus properties occurring earlier on, and more abstract semantic and categorical  
381 responses occurring later (Carlson, Simmons, Kriegeskorte, & Slevc, 2014; Carlson, et al., 2013;  
382 Cichy, et al., 2014; Clarke & Tyler, 2014). It should also be noted that the models show high  
383 correlations even though our design incorporated left-right flips of the images and spatial  
384 displacement of the images to reduce the influence of low-level stimulus properties. The low-level  
385 models were generated using only the standard orientation of each stimulus at a fixed position, yet  
386 could still predict the data after these transformations, affirming the importance of low-level visual  
387 similarity in the initial representation of the stimuli.

388 **Contemporary models: Intermediate processing emphasizes faces and bodies**

389 A wide range of theoretical models have been proposed to account for the brain's higher-order  
390 representation of objects. We tested how each of these models could account for the brain's  
391 emerging representation of the objects (Figure 3B-D). The models we tested included a range of  
392 categorical models (e.g., animate versus inanimate), as well as a biological continuum model (Sha  
393 et al., 2015). We assessed their explanatory power using RSA and found that the models produced  
394 varying results. Starting at approximately 100ms, the face/body category model had the most  
395 explanatory power. Notably, the two other models with the early peaks (animal-cluster and  
396 faces/bodies) were among those showing significant overlap with one or both of the low-level  
397 feature models (see Figure 2E), suggesting that low-level visual similarity may have contributed to  
398 the earlier onset and high peaks for these models. At approximately 300ms, there appeared to be a  
399 transition in the representational structure. Here, the faces/body model, which was the best  
400 performing model in the early period (100-200ms), has declined and the biological continuum  
401 model increased its performance to have comparable explanatory power to the faces/bodies model.  
402 Interestingly, the animacy model was among the weakest performing models, despite a number of  
403 studies showing animacy provides a significant account of the human and primate brain's higher-  
404 order representation of objects (e.g., Carlson et al., 2013; Cichy et al., 2014; Kiani et al., 2007;  
405 Kriegeskorte et al., 2008b).

406

407 **Human similarity model provides the best account of late representations**

408 Higher order factors such as human similarity and agency are known to influence human perception  
409 of objects (Gobbini et al., 2011; Gray et al., 2007). To assess these attributes, we collected  
410 behavioural ratings for the stimuli about various higher order attributes (e.g., capacity to experience  
411 pleasure), to generate a new set of models. We then tested whether these models could account for  
412 the brain's emerging representation of objects (Figure 3D). Across the behavioural-rating models,  
413 the results were very similar, which can be attributed to the high level of overlap in their internal

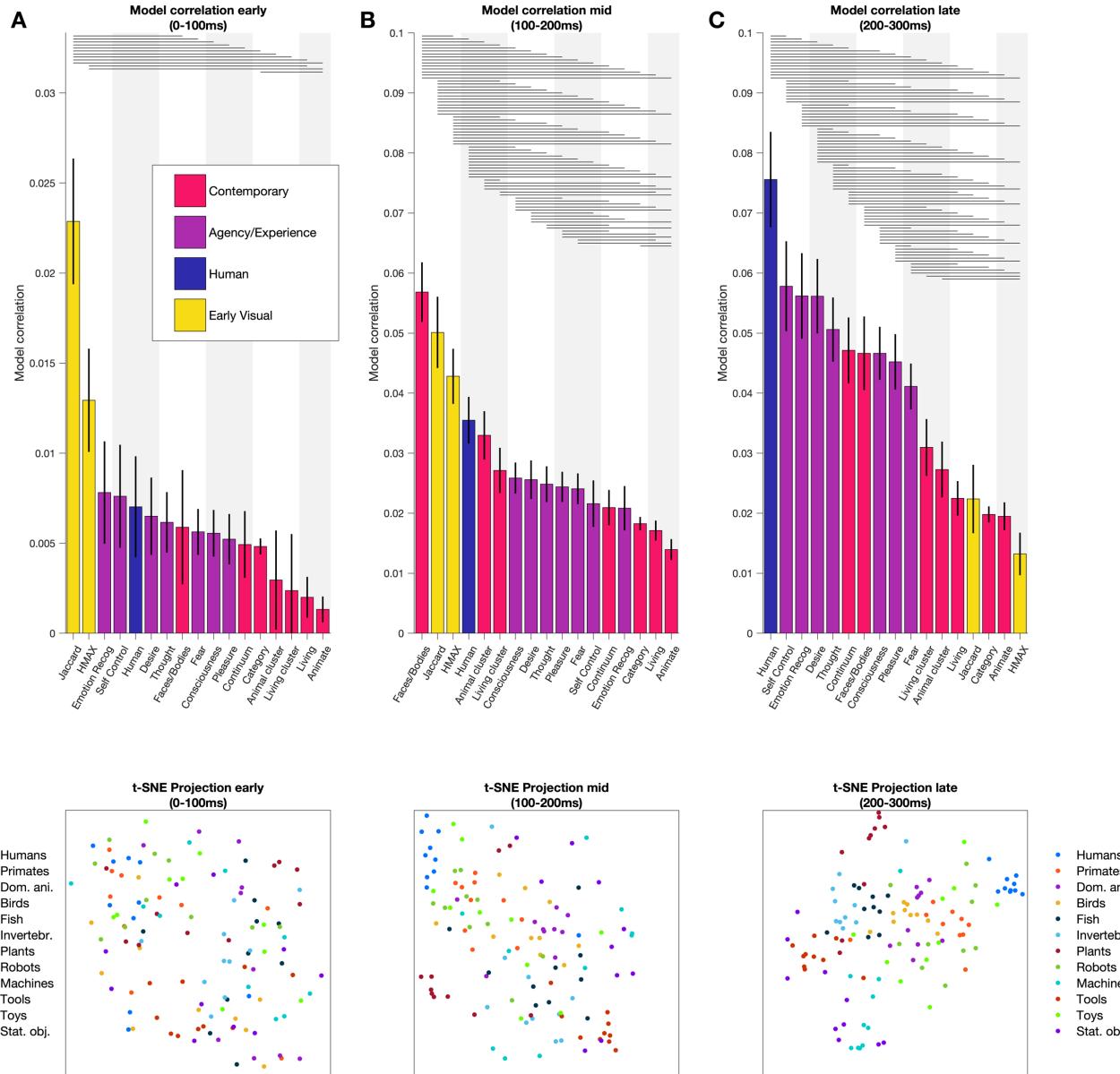
414 structure (see Figure 2F). The models all show an initial peak at approximately 90ms, and then rise  
415 to a more significant peak at about 245 to 280 ms. This pattern was also observed for the biological  
416 continuum model, as seen in (Figure 3C). Notably, the best performing model in the later time  
417 window is the human similarity model, which is based on the question “How similar is this (object)  
418 to a human?”

419

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

421 Our analysis of the models’ performance in the time series broadly indicated three distinct stages of  
422 processing. Early in the time series (<100ms) the low-level feature models performed the best.  
423 Next, in the intermediate time from 100ms to 200ms, three category models (Face/Body, Animal  
424 cluster, and Living Cluster) exhibited a distinct peak in their performance at about 180ms. Finally,  
425 the higher-order models based on agency and experience, and the biological continuum model,  
426 showed a slow rise that peaked about 270ms. To quantify these observations, we discretised the  
427 data in three 100ms time windows (0-100ms; 100-200ms; 200-300ms) in which we compared all  
428 model performances. Figure 4 shows the results of the windowed RSA analysis. Below each plot  
429 are visualisations of the representations at each stage constructed by projecting the data into two  
430 dimensions using t-SNE (Maaten & Hinton, 2008).

431 To compare the models, we conducted a series of t-tests between all pairwise models to  
432 assess between-model performance in each of these time windows separately (adjusted for multiple  
433 comparisons across time points using an FDR of  $q < .01$ ). In the early time window (0-100ms), the  
434 Jaccard (shape) model had the highest correlation (Figure 4A). In the second time window (100-  
435 200ms), the low-level models (Jaccard/ HMAX) still performed well; however, the face/body  
436 category model was the best performing model (Figure 4B). In the final time window (200-300ms),  
437 the low-level models (Jaccard/ HMAX) were among the worst performing models; and the human  
438 similarity model was the best model overall (Figure 4C). Notably, the humanness model  
439 outperformed the faces/bodies model, suggesting that the organisation of neural representations



440

441 **Figure 4.** Model correlations (Kendall's tau-a) in the early (0-100ms; A), middle (100-200ms; B)  
442 and late (200-300ms; C) time windows. Models are arranged in order of highest average rank  
443 correlation within each time window (highest on the left), and error bars indicate 95% SEM of the  
444 model correlation. Paired t-tests determined significant differences between model correlations  
445 (lines indicate significance at  $p < .05$ , adjusted for multiple comparisons across time points using a  
446 FDR of  $q < .01$ ). Note the y-axis range for the early window is different because the model  
447 correlations were lower. Below each plot, two-dimensional embeddings of the stimuli are shown.  
448 The embeddings were computed using t-SNE (Maaten and Hinton, 2008). The distance between  
449 two points in this embedding reflects their neural dissimilarity.

450

451

452

453 according to this human-centric model was not merely driven by the presence of faces/bodies,  
454 which are well-established as significant factors in object processing (Farah, 1996; Kanwisher et al.,  
455 1997; Thorpe et al., 1996). In terms of the model ranking, it can also be seen that the all the  
456 agency/experience models have good fits with the data, performing close or better than to the  
457 face/body category model and continuum model.

458 Overall, the results of the windowed analysis conforms with the observations from the time  
459 series data. Low level feature models provided the best fit for early representations, category  
460 models (in particular face/body, and the animal and living cluster) and low level feature models  
461 explained intermediate representations, and higher order models best explained late stage object  
462 representations. And, at this late stage of processing, the human similarity model provided the best  
463 fit to the data overall.

464

## 465 Discussion

466 Many classifications, such as animate/inanimate and living/non-living, have been proposed as  
467 organisational principles for the brain's representation of objects. Here we sought to provide an in-  
468 depth evaluation of contemporary models of visual object representations by evaluating their  
469 capacity to account for neural responses to a diverse range of object stimuli. In addition to these  
470 contemporary models and models based on low-level visual similarity, we created new theoretical  
471 and behaviour-based models. To test the predictive power of these models, we included novel  
472 stimuli that did not conform to the typical categories, such as robots and toys. Our results showed  
473 that the best performing model overall for late stage processing of objects was one based on the  
474 broad concept of human-similarity.

475 Our findings are consistent with accepted knowledge about the flow of information in  
476 human object recognition (for review see Contini, Wardle, & Carlson, 2017). This multi-stage  
477 processes begins with processing low-level visual properties of the stimulus, presumably in early  
478 visual cortex. These early representations are then subsequently transformed into higher order

479 representations incorporating category structure (Carlson, et al., 2013; Cichy, Khosla, Pantazis,  
480 Torralba, & Oliva, 2016; Contini, et al., 2017) and semantic information (Carlson, et al., 2014;  
481 Clarke & Tyler, 2014). We found the best performing models at early time points were the low-  
482 level feature models (Jaccard and HMAX), and higher order models fit the data better later in the  
483 time series.

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

503 The superior performance of the human model builds on our existing understanding of the  
504 representation of object categories in the brain. The model extends on the continuum idea of

505 Connolly et al. (2012) and Sha et al. (2015), as it represents a type of human-similarity continuum  
506 (see also Thorat, et al., 2019). However, unlike the animacy continuum that is based on biological  
507 classes, the human model was not limited by biology (Gobbini, et al., 2011; Tong, Nakayama,  
508 Moscovitch, Weinrib, & Kanwisher, 2000). Results from an fMRI study by Gobbini et al. (2011)  
509 are also consistent with a level of cross-over between animate/inanimate object categories that does  
510 not fit into this dichotomy, nor a continuum based on biological classes. The authors compared  
511 human observers' perception of human faces and robots and found that robots evoked activation in  
512 areas associated with faces (though to a lesser extent than humans), while also activating object  
513 areas and areas associated with mechanical movements. This supports the idea of more a complex  
514 model of object categorisation that incorporates factors such as agency and human-related  
515 experiences. Given the relative strength of our human-centric model in accounting for the data, the  
516 idea of "humanness" as an important dimension in the neural representation of objects warrants  
517 further exploration.

518 Our best performing human-centric model likely encompasses a complex set of features,  
519 including both visual and conceptual factors. In our study, we did not impose a definition or any  
520 criteria against which people should rate the objects when asked 'How similar is it to a human?'  
521 (with responses from this survey used to generate the human model). Accordingly, we do not know  
522 which features people were using to rate object 'humanness', raising an interesting area for further  
523 investigation. The brain likely makes use of both visual and semantic information for representing  
524 objects (Carlson, et al., 2014; Clarke & Tyler, 2014; Coggan, Baker, & Andrews, 2016). Our data  
525 suggests that the semantic component of object representations incorporates information about  
526 concepts such as function, agency, and human experience. Indeed, a recent study by Connolly et al.  
527 (2016) showed an overlap between regions sensitive to the perceived threat of animals and those  
528 associated with social cognition, highlighting the importance of agent-related dimensions to object  
529 processing.

530 Presently, we still do not have a clear understanding of how different semantic concepts  
531 relate to object representations and category structure. A recent model attempts to explain the neural  
532 representation of object attempts using a multidimensional framework (Martin, 2016). In this paper,  
533 the author suggests that neural patterns associated with objects are formed from complex interactive  
534 circuits based on a range of systems throughout the brain, including those associated with action,  
535 perception and emotion. This idea shifts the focus away from models based on categories, with a  
536 view to a more holistic approach to object representations that considers interactions between  
537 various circuits throughout the brain. In this multidimensional framework, it is essential to  
538 recognize that no single feature or attribute could be able to fully explain the richness of the brain's  
539 (multidimensional) representation of objects (see e.g., Thorat, et al., 2019). Recent fMRI studies  
540 have sought to identify principle axes of object representations in the brain (e.g., Connolly, et al.,  
541 2012; Sha, et al., 2015; Thorat, et al., 2019). In the present study, we show that the human  
542 similarity provided the best account of late stage processing, highlighting "humanness" as a key  
543 feature in the human brain's representation of objects that shapes our experience of the world.  
544

545 **Acknowledgements:**

546 This research was supported by an Australian Research Council Future Fellowship (FT120100816)  
547 and an Australian Research Council Discovery project (DP160101300) awarded to T.A.C. The  
548 authors acknowledge the University of Sydney HPC service for providing High Performance  
549 Computing resources. The authors declare no competing financial interests.  
550

551 **Author contributions:**

552 E.C, E.G, M.W, T.C designed the study  
553 E.C, T.G collected the data  
554 E.C, E.G, T.G conducted the analysis  
555 E.C, E.G, T.C interpreted the results  
556 E.C and T.C wrote the manuscript  
557 All authors reviewed the manuscript  
558

## 559 References

- 560 Barragan-Jason, G., Cauchoux, M., & Barbeau, E. J. (2015). The neural speed of familiar face  
561 recognition. *Neuropsychologia*, 75, 390-401.
- 562 Bracci, S., Kalfas, I., & Op de Beeck, H. (2017). The ventral visual pathway represents animal  
563 appearance over animacy, unlike human behavior and deep neural networks. *BioRxiv*,  
564 228932.
- 565 Bracci, S., & Op de Beeck, H. (2016). Dissociations and Associations between Shape and Category  
566 Representations in the Two Visual Pathways. *J Neurosci*, 36, 432-444.
- 567 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, 10, 433-436.
- 568 Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the  
569 animate-inanimate distinction. *J Cogn Neurosci*, 10, 1-34.
- 570 Carlson, T. A., Goddard, E., Kaplan, D. M., Klein, C., & Ritchie, J. B. (2018). Ghosts in machine  
571 learning for cognitive neuroscience: Moving from data to theory. *Neuroimage*, 180, 88-100.
- 572 Carlson, T. A., Simmons, R. A., Kriegeskorte, N., & Slevc, L. R. (2014). The emergence of semantic  
573 meaning in the ventral temporal pathway. *J Cogn Neurosci*, 26, 120-131.
- 574 Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of  
575 object vision: the first 1000 ms. *J Vis*, 13.
- 576 Cauchoux, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The neural dynamics of face  
577 detection in the wild revealed by MVPA. *J Neurosci*, 34, 846-854.
- 578 Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex  
579 for perceiving and knowing about objects. *Nat Neurosci*, 2, 913-919.
- 580 Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal  
581 stream. *Neuroimage*, 12, 478-484.
- 582 Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural  
583 networks to spatio-temporal cortical dynamics of human visual object recognition reveals  
584 hierarchical correspondence. *Sci Rep*, 6, 27755.
- 585 Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and  
586 time. *Nat Neurosci*, 17, 455-462.
- 587 Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *J  
588 Neurosci*, 34, 4766-4775.
- 589 Coggan, D. D., Baker, D. H., & Andrews, T. J. (2016). The Role of Visual and Semantic Properties in  
590 the Emergence of Category-Specific Patterns of Neural Response in the Human Brain.  
591 *eNeuro*, 3.
- 592 Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., Abdi, H., & Haxby,  
593 J. V. (2012). The representation of biological classes in the human brain. *J Neurosci*, 32,  
594 2608-2618.
- 595 Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object  
596 recognition in the human brain: From visual features to categorical decisions.  
597 *Neuropsychologia*, 105, 165-176.
- 598 Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain  
599 specificity in visual cortex. *Cereb Cortex*, 16, 1453-1461.
- 600 Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual  
601 processing of the human body. *Science*, 293, 2470-2473.
- 602 Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area:  
603 recognition, navigation, or encoding? *Neuron*, 23, 115-125.
- 604 Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment.  
605 *Nature*, 392, 598-601.

- 606 Gainotti, G. (2000). What the locus of brain lesion tells us about the nature of the cognitive defect  
607 underlying category-specific disorders: a review. *Cortex*, 36, 539-559.
- 608 Gobbini, M. I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., Guazzelli, M., & Pietrini, P.  
609 (2011). Distinct neural systems involved in agency and animacy detection. *J Cogn Neurosci*,  
610 23, 1911-1920.
- 611 Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016). Representational dynamics of  
612 object recognition: Feedforward and feedback information flows. *Neuroimage*, 128, 385-  
613 397.
- 614 Goddard, E., Klein, C., Solomon, S. G., Hogendoorn, H., & Carlson, T. A. (2018). Interpreting the  
615 dimensions of neural feature representations revealed by dimensionality reduction.  
616 *Neuroimage*, 180, 41-67.
- 617 Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315, 619.
- 618 Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from  
619 Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series  
620 Neuroimaging Data. *J Cogn Neurosci*, 29, 677-697.
- 621 Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed  
622 and overlapping representations of faces and objects in ventral temporal cortex. *Science*,  
623 293, 2425-2430.
- 624 Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The  
625 functional organization of human extrastriate cortex: a PET-rCBF study of selective  
626 attention to faces and locations. *J Neurosci*, 14, 6336-6353.
- 627 Haynes, J. D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and  
628 Perspectives. *Neuron*, 87, 257-270.
- 629 Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes  
630 the representation of thousands of object and action categories across the human brain.  
631 *Neuron*, 76, 1210-1224.
- 632 Jordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2016). Typicality sharpens category  
633 representations in object-selective cortex. *Neuroimage*, 134, 170-179.
- 634 Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed  
635 representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A*,  
636 96, 9379-9384.
- 637 Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des  
638 Jura. *Bulletin de la Société vaudoise des sciences naturelles*, 37.
- 639 Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human  
640 extrastriate cortex specialized for face perception. *J Neurosci*, 17, 4302-4311.
- 641 Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response  
642 patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol*, 97,  
643 4296-4309.
- 644 Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36, 14-14.
- 645 Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and  
646 object size. *J Neurosci*, 33, 10235-10242.
- 647 Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition,  
648 computation, and the brain. *Trends Cogn Sci*, 17, 401-412.
- 649 Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting  
650 the branches of systems neuroscience. *Front Syst Neurosci*, 2, 4.
- 651 Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P.  
652 A. (2008). Matching categorical object representations in inferior temporal cortex of man  
653 and monkey. *Neuron*, 60, 1126-1141.

- 654 Maaten, L. v. d., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning  
655 Research*, 9, 2579–2605.
- 656 Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumia, R. I., Caramazza, A., & Martin, A. (2007). Action-  
657 related properties shape object representations in the ventral stream. *Neuron*, 55, 507-  
658 520.
- 659 Martin, A. (2016). GRAPES-Grounding representations in action, perception, and emotion systems:  
660 How object properties and categories are represented in the human brain. *Psychon Bull  
661 Rev*, 23, 979-990.
- 662 Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox  
663 for representational similarity analysis. *PLoS Comput Biol*, 10, e1003553.
- 664 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into  
665 movies. *Spat Vis*, 10, 437-442.
- 666 Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial  
667 overview. *Neuroimage*, 45, S199-209.
- 668 Pinski, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S. (2009).  
669 Neural representations of faces and body parts in macaque and human cortex: a  
670 comparative fMRI study. *J Neurophysiol*, 101, 2581-2600.
- 671 Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and  
672 Object Category in Human Visual Cortex: The Animate-Inanimate Distinction. *J Cogn  
673 Neurosci*, 28, 680-692.
- 674 Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human  
675 visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging  
676 study. *J Neurosci*, 16, 5205-5215.
- 677 Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat  
678 Neurosci*, 2, 1019-1025.
- 679 Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object  
680 processing. A positron emission tomography study. *Brain*, 115 Pt 1, 15-36.
- 681 Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory  
682 of immediate visual recognition. *Prog Brain Res*, 165, 33-56.
- 683 Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C.  
684 (2015). The animacy continuum in the human ventral vision pathway. *J Cogn Neurosci*, 27,  
685 665-678.
- 686 Taylor, J. C., & Downing, P. E. (2011). Division of labor between lateral and ventral extrastriate  
687 representations of faces, bodies, and objects. *J Cogn Neurosci*, 23, 4122-4137.
- 688 Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human  
689 ventral temporal cortex. *arxiv.org, arXiv:1904.02866*.
- 690 Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*,  
691 381, 520-522.
- 692 Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties  
693 of the human fusiform face area. *Cogn Neuropsychol*, 17, 257-280.
- 694 van de Nieuwenhuijzen, M. E., Backus, A. R., Bahramisharif, A., Doeller, C. F., Jensen, O., & van  
695 Gerven, M. A. (2013). MEG-based decoding of the spatiotemporal dynamics of visual  
696 category perception. *Neuroimage*, 83, 1063-1073.
- 697 Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107 ( Pt 3),  
698 829-854.
- 699