

27 **Abstract**

28 Animate and inanimate objects elicit distinct response patterns in the human ventral
29 temporal cortex (VTC), but the exact features driving this distinction are still poorly
30 understood. One prominent feature that distinguishes typical animals from inanimate
31 objects and that could potentially explain the animate-inanimate distinction in the VTC is
32 the presence of a face. In the current fMRI study, we investigated this possibility by
33 creating a stimulus set that included animals with faces, faceless animals, and inanimate
34 objects, carefully matched in order to minimize other visual differences. We used both
35 searchlight-based and ROI-based representational similarity analysis (RSA) to test
36 whether the presence of a face explains the animate-inanimate distinction in the VTC.
37 The searchlight analysis revealed that when animals with faces were removed from the
38 analysis, the animate-inanimate distinction almost disappeared. The ROI-based RSA
39 revealed a similar pattern of results, but also showed that, even in the absence of faces,
40 information about agency (a combination of animal's ability to move and think) is present
41 in parts of the VTC that are sensitive to animacy. Together, these analyses showed that
42 animals with faces do elicit a stronger animate/inanimate response in the VTC, but that
43 this effect is driven not by faces per se, or the visual features of faces, but by other factors
44 that correlate with face presence, such as the capacity for self-movement and thought.
45 In short, the VTC appears to treat the face as a proxy for agency, a ubiquitous feature of
46 familiar animals.

47

48 **Key words:**

49 fMRI, ventral temporal cortex, object representations, face perception, animacy

50

51 **Significance Statement**

52 Many studies have shown that images of animals are processed differently from
53 inanimate objects in the human brain, particularly in the ventral temporal cortex (VTC).
54 However, what features drive this distinction remains unclear. One important feature that
55 distinguishes many animals from inanimate objects is a face. Here, we used fMRI to test
56 whether the animate/inanimate distinction is driven by the presence of faces. We found
57 that the presence of faces did indeed boost activity related to animacy in the VTC. A more

58 detailed analysis, however, revealed that it was the association between faces and other
59 attributes such as the capacity for self-movement and thinking, not the faces *per se*, that
60 was driving the activity we observed.

61

62 Introduction

63 Multiple studies have shown that an animate/inanimate distinction is a major factor in the
64 organization of object representations in the human ventral temporal cortex (VTC)
65 (Kriegeskorte et al., 2008; Bracci and Op de Beeck, 2016; Proklova et al., 2016). Animacy,
66 however, is associated with different attributes, from visual features to semantic concepts
67 such as agency (Peelen and Downing, 2017), and it is still unclear what drives the
68 animate/inanimate distinction in the VTC. Some studies have suggested that the it can
69 be explained by visual features alone (Long et al., 2018; Coggan et al., 2016). Others
70 have suggested that the animate/inanimate distinction is not based purely on visual
71 features but is also driven by conceptual information associated with animacy (Thorat et
72 al., 2019; Proklova et al., 2016; Sha et al., 2016).

73 Stimuli that do not conform to the strict animate/inanimate dichotomy have recently
74 provided new insights into animacy organization in the VTC. A recent fMRI study showed
75 that inanimate objects that share features with animals (e.g., cow-shaped mugs) are
76 represented in a similar way to animate objects (Bracci et al., 2018). Another study that
77 used “borderline” stimuli, such as robots, together with more typical animate or inanimate
78 objects, found similar results using MEG (Contini et al., 2019). This suggests that what
79 matters for the VTC is similarity to humans (either visual or semantic), rather than animacy
80 *per se* (Contini et al., 2019; Sha et al., 2016; Gobbini et al., 2011; Thorat et al., 2019).
81 Another possibility is that the VTC is tuned to the diagnostic features of animacy shared
82 between animals and animal-like stimuli (e.g., faces or bodies; Bracci et al., 2018).

83 Although there is evidence that body shape cannot fully account for the
84 representation of animacy in the VTC (Proklova et al., 2016, Bracci et al., 2016), few
85 previous studies have controlled for the presence of a face. Faces are extremely
86 biologically relevant stimuli. They play an important role in determining whether
87 something is animate – and the animals with which people are most familiar tend to have
88 faces. Animal-like objects such as robots and toys often share this critical feature with
89 animals, which could explain why they are represented similarly to animals in the VTC.

90 A network of areas mediates face perception, most notably the fusiform face area
91 (FFA) in the VTC (Grill-Spector et al., 2018). Interestingly, in a study by Proklova et al.
92 (2016), the fusiform gyrus was one of the areas in which animacy information was present

93 after controlling for most visual features. Since all the animate stimuli in that study had
94 faces, it is possible that this effect was driven by the presence of faces rather than by
95 animacy *per se*.

96 A number of neuroimaging studies have argued that faces are not necessary for
97 eliciting typical animate/inanimate distinction in the VTC by using images of animals with
98 their faces covered (Chao et al., 1999), geometric shapes moving in characteristically
99 animate (or inanimate) ways (Martin and Weisberg, 2004), or synthetic stimuli (*texforms*)
100 which preserve the mid-level features of the stimuli but not the fine detail such as faces
101 (Long et al., 2018). To our knowledge, however, no previous studies have directly
102 compared the patterns of brain activation generated by images of real animals with and
103 without faces. In the present fMRI study, we addressed this gap by using a stimulus set
104 that included images of real animals with and without a face, as well as inanimate objects,
105 all of which had otherwise similar visual features. Thus, we made animacy orthogonal to
106 the presence of a face, which allowed us to directly examine the role of faces in VTC
107 representations. We found that when animals with faces were removed from the analysis,
108 the animate/inanimate distinction largely disappeared. Nevertheless, additional analyses
109 revealed that this effect was driven not by faces *per se*, but by other features that typically
110 correlate with the presence of a face, such as mobility and thoughtfulness.

111

112 **Materials and methods**

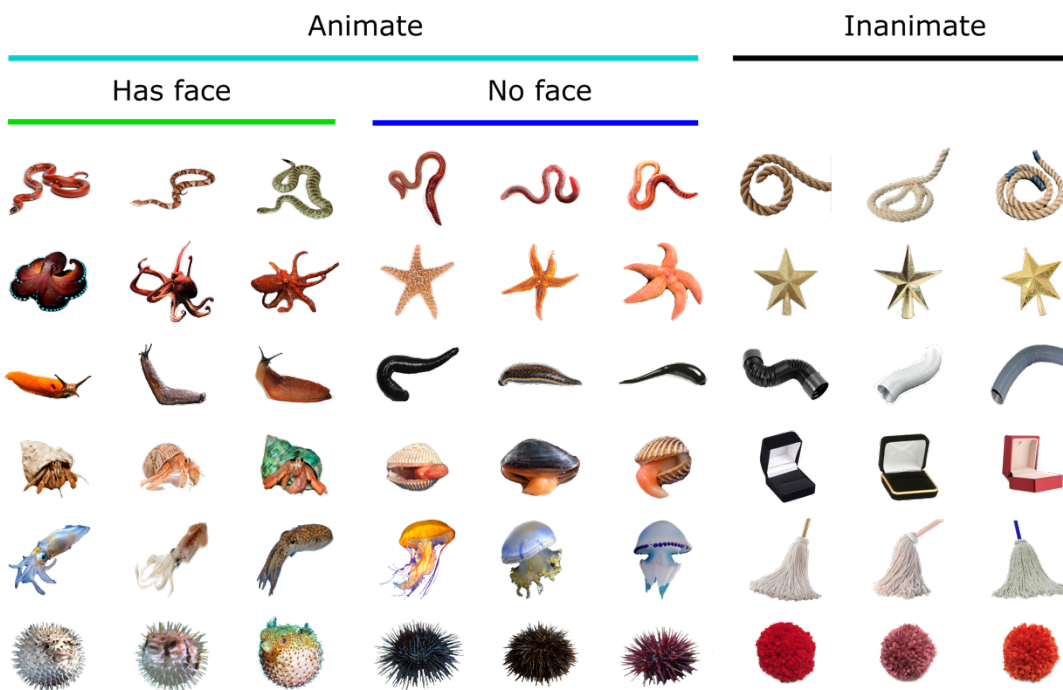
113 *Participants*

114 Twenty-four volunteers (18 female, mean age 26.7 years, SD = 3.3) participated in the
115 behavioral ratings experiment. These included 14 people who also took part in the fMRI
116 experiment described below (rating the animals after the fMRI session) and ten who did
117 not. Ten other volunteers, who were not part of the behavioral ratings or the fMRI
118 experiments (4 female, mean age = 25 years, SD = 4.6) participated in the behavioral
119 visual search experiment. Twenty volunteers (14 female, mean age 26.1 years, SD = 4)
120 took part in the fMRI study. Two were excluded from further analysis because of
121 excessive head movement. All participants gave informed consent and the protocol was
122 approved by the University of Western Ontario Ethics Review Board.

123

124 *Stimuli*

125 The stimulus set (Fig. 1) consisted of 18 unique stimuli divided into 3 groups: 6 animals
126 with distinct faces (e.g., snake), 6 animals without a distinct face (e.g., starfish), and 6
127 inanimate objects. Additionally, we used three different exemplars of each stimulus,
128 resulting in a total of 54 stimuli. Importantly, to minimize the effects of visual feature
129 similarity, stimuli were organized into six triplets (animal with a face - faceless animal -
130 object) based on the overall shape similarity, such that all three stimuli within each triplet
131 shared a similar overall shape (e.g., snake – worm – rope).



132
133 **Figure 1. Stimulus set.** The stimulus set included images of animals with faces (left 3
134 columns), without faces (middle 3 columns), and inanimate objects (right 3 columns).
135 There were three exemplars of each unique animal/object. To minimize visual differences
136 between the stimuli from different categories, the stimuli in each row shared a similar
137 shape.

138
139 *Behavioral ratings experiment*

140 Participants were asked to rate the 12 animal stimuli presented on a set of 6
141 characteristics: (1) Does this animal have a face? (2) How fast can this animal move? (3)
142 Is this animal capable of thoughts? (4) Does this animal have a head? (5) Does this animal

143 have eyes? (6) How familiar is this animal to you? There was a separate block for each
144 of the 6 questions. The question appeared at the beginning of the block, followed by a
145 presentation of all twelve animals used in the experiment. On each trial, participants were
146 presented with all 3 versions of a given animal and the continuous ratings bar on the
147 computer screen (Fig. 2C). Using a mouse, participants could click anywhere on a bar to
148 provide their response, ranging from 0 (e.g. “this animal is not at all familiar”) to 100 (“I
149 am very familiar with this animal”). In this and all the following experiments, including
150 fMRI, stimuli were presented using Psychtoolbox for Matlab (Brainard, 1997).

151

152 *Behavioral visual search experiment*

153 The visual search experiment was analogous to the one described in Proklova et al.
154 (2016) and based on the approach by Mohan and Arun (2012). On each trial, participants
155 saw an array of 16 stimuli, 15 of which were identical distractors and one was an oddball
156 target (see Fig. 2B for a trial example). We measured participants’ reaction times (RTs)
157 to find the target and used the RT as a proxy of visual similarity between target and
158 distractor. Faster RTs suggest that target “popped out” more because it was more visually
159 distinct from distractors, and slower RTs indicate that the search was harder because the
160 target was quite similar visually to the distractors. Each of the stimuli (Fig. 1) appeared
161 both as a target and a distractor during the experiment, giving us the visual similarity
162 estimate for each pair of stimuli. This excluded comparisons between the 3 versions of
163 the same animal or object, since the responses for 3 versions were averaged in the fMRI
164 experiment and were not analyzed separately in the representational similarity analysis.
165 We refer the reader to Proklova et al. (2016) for further details of the visual search
166 experiment.

167

168 *fMRI acquisition*

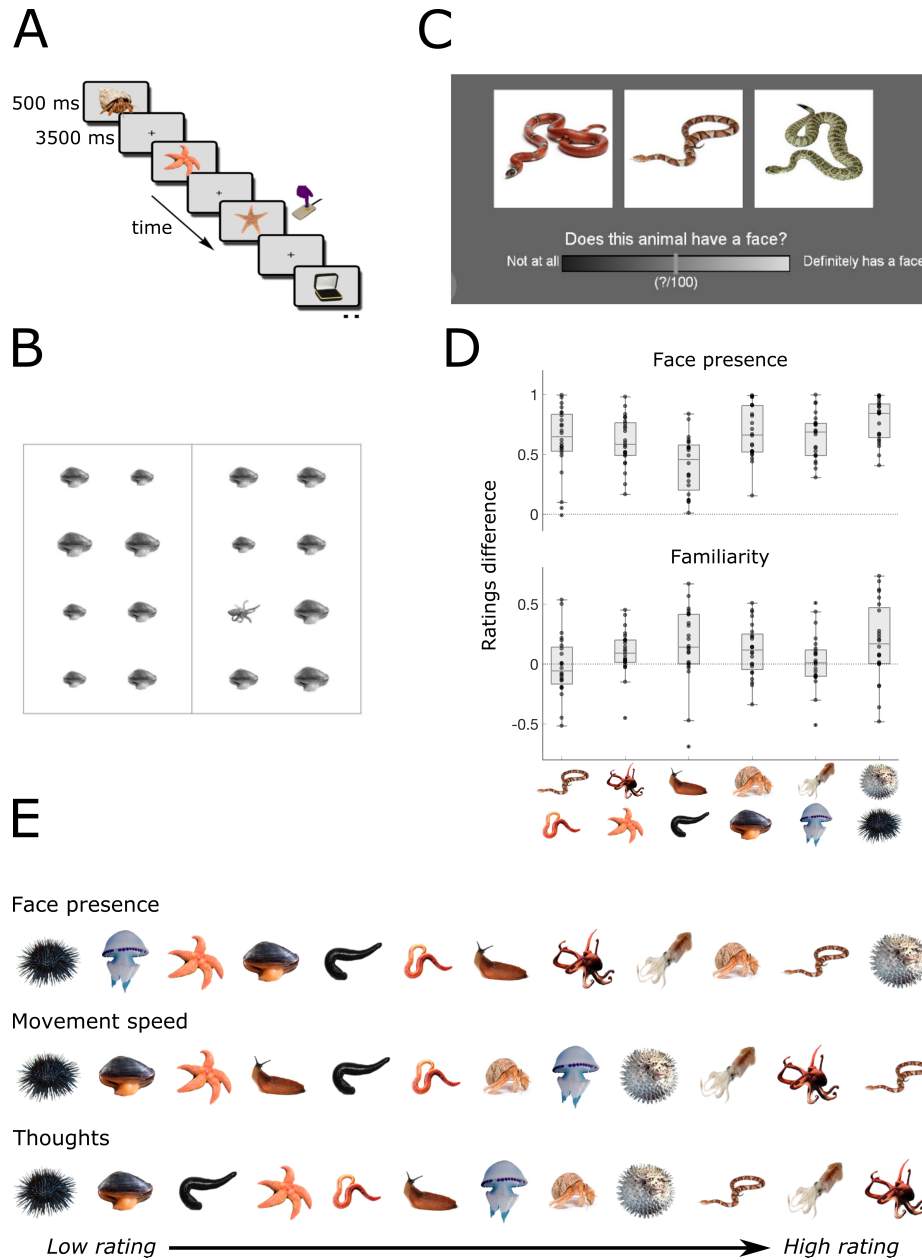
169 fMRI data were acquired using a Siemens MAGNETOM Prisma Fit whole-body 3 Tesla
170 MRI scanner with a 32-channel head coil at the Centre for Functional and Metabolic
171 Mapping, Western University, London, Ontario. The stimuli were back projected onto a
172 translucent screen placed inside the scanner bore (60 Hz refresh rate; 1024 × 768 spatial
173 resolution). Participants viewed the stimuli via a mirror mounted to the head coil.

174 Functional images were collected with an echo planar imaging sequence (echo time, 30
175 ms; repetition time, 2000 ms; field of view, $196 \times 196 \text{ mm}^2$; matrix, 64×64 ; flip angle,
176 90° ; slice thickness, 3 mm; gap, 0 mm; number of slices, 36; axial slice orientation). A
177 high-resolution 3D structural T1-weighted scan was collected at the beginning of the
178 session using a magnetization-prepared rapid gradient echo sequence (MPRAGE) with
179 voxel size of 1 mm isotropic.

180

181 *fMRI task*

182 While in the scanner, participants completed 8 runs of the main experiment, and 2
183 functional localizer runs. In each functional run, each of the 54 stimuli appeared at least
184 once, in randomized order. Some stimuli appeared twice in order to introduce one-back
185 repetition trials (10 per block.) Participants' task was to press the button any time the
186 image of the same animal or object (e.g., 2 snakes, but not necessarily the exact same
187 photo of a snake) appeared on 2 consecutive trials. The button-press trials were not
188 further analyzed. Stimuli were on screen for 500 ms, followed by 3500-ms inter-stimulus
189 interval (Fig. 2A). Each run lasted 256 s. In the functional localizer runs, participants saw
190 images from 4 categories (faces, bodies, animals, objects) in a blocked design, pressing
191 the button when the same image appeared twice in a row. The localizer procedure is
192 described in detail in Proklova et al. (2016).



193
194 **Figure 2. Experimental design and behavioral ratings analysis.** **A**, fMRI experimental
195 design. Stimuli appeared one at a time, and participants were instructed to press a button
196 whenever the same animal or object appeared on 2 consecutive trials. **B**, Trial example
197 from the behavioral visual search experiment used for quantifying pairwise visual
198 dissimilarity between the stimuli. Participants had to find an oddball image among the
199 distractors as quickly as possible and indicate with a button press whether it was on the
200 left or right of the central line. **C**, Trial example in a behavioral ratings experiment.
201 Participants had to indicate their response by clicking the mouse at any point of a
202 continuous ratings bar. **D**, The difference between the face presence (upper panel) and
203 familiarity (lower panel) ratings for each of the 6 animal pairs. Above zero values indicate
204 that, for the 2 animals sharing a similar shape, the animal with a face received higher

205 rating than an animal without a distinct face. **E**, Animate stimuli sorted based on group-
206 averaged behavioral ratings, from the lowest to the highest.

207

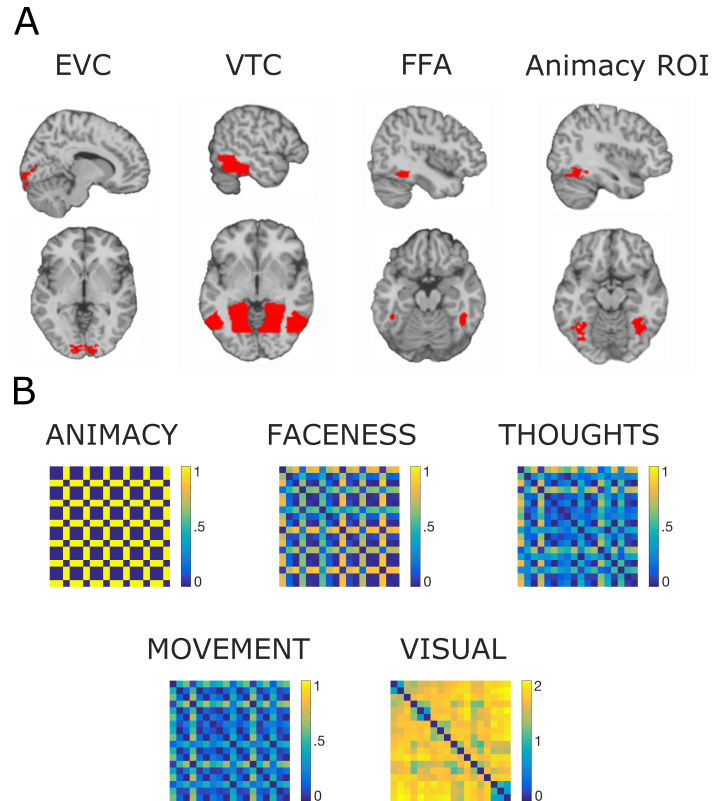
208 *fMRI preprocessing*

209 The neuroimaging data were analyzed using Matlab and SPM12. Preprocessing involved
210 realigning the functional volumes, coregistering them to the structural image, re-sampling
211 to a $2 \times 2 \times 2$ mm grid, and spatially normalizing to the Montreal Neurological Institute
212 305 template included in SPM12. For the univariate analysis of the localizer data, the
213 functional images were smoothed with a 6-mm FWHM kernel. Images were not smoothed
214 for the multivariate analyses of the main experiment data. For the main experiment, the
215 BOLD signal of each voxel in each participant was modeled using 24 regressors in a
216 general linear model, with 18 regressors for each of the objects (e.g., one regressor for
217 all snakes) and six regressors for the movement parameters obtained from the
218 realignment procedure.

219 *ROI definition*

220 The four regions of interest (ROIs) are shown in Figure 3A and were defined as follows.
221 Early visual cortex (EVC) ROI was defined anatomically by selecting Brodmann area 17
222 (BA17) using WFU PickAtlas toolbox for SPM12 (Maldjian et al., 2003), and spanned
223 5,856 mm³ in size. To define fusiform face area (FFA), we used the group-level Faces >
224 Houses contrast in the functional localizer ($p < 0.05$, FWE corrected), which revealed two
225 clusters: one in the right hemisphere (792 mm³, peak MNI coordinates: $x = 46$, $y = -50$, z
226 $= -20$), and one in the left hemisphere (208 mm³, peak MNI coordinates: $x = -42$, $y = -40$,
227 $z = -20$). Ventral temporal cortex (VTC) was defined according to previous studies (Haxby
228 et al., 2011, Thorat et al., 2019). It included the inferior temporal, fusiform, and
229 lingual/parahippocampal gyri, and extended from -71 to -21 on the y -axis of the MNI
230 coordinates, with total volume of 73,776 mm³. Finally, we included a region defined in an
231 earlier study (Proklova et al., 2016) in which the animate/inanimate distinction was
232 observed after controlling for visual features (this study involved different stimuli and
233 participants.) We refer to it as “Animacy ROI” for simplicity. It consisted of two clusters,
234 with peaks in right fusiform gyrus ($x = 42$, $y = -60$, $z = -18$; 2,672 mm³) and left fusiform
235 gyrus ($x = -44$, $y = -52$, $z = -16$; 1,968 mm³). The specific analyses that were used to

236 define this region are described in Proklova et al. (2016). The Animacy ROI and the FFA
237 only shared 10 voxels (80 mm³) in common, and both were almost fully encompassed by
238 the VTC: 97% of FFA (122/125 voxels) and 83% of the Animacy ROI (482/580 voxels)
239 intersected with the larger VTC ROI.



240

241 **Figure 3. Regions of interest and model RDMs.** **A**, Regions of Interest included early
242 visual cortex, ventral temporal cortex, fusiform face area, and the region in which the
243 animate/inanimate distinction was observed after controlling for visual features in an
244 earlier study (Animacy ROI). **B**, Model representational dissimilarity matrices (RDMs)
245 used in the representational similarity analysis.

246

247 *Representational Similarity Analysis (RSA) Searchlight*

248 The RSA searchlight procedure was analogous to the one used in previous studies
249 (Proklova et al, 2016; Thorat et al., 2020). All analyses were done using CoSMoMVPA
250 toolbox for Matlab (Oosterhof et al., 2016). For each voxel in the brain, we took a 100-
251 voxel spherical neighborhood around it and measured voxel-wise response patterns for
252 all 18 conditions in this region. We then calculated pairwise Pearson correlations between
253 these patterns for all pairs of stimuli. These correlations were then inversed (1-Pearson)
254 and used to create a 18 x 18 neural representational dissimilarity matrix (RDM), in which

255 every entry corresponded to the dissimilarity between a pair of stimuli. For each 100-voxel
256 neighborhood, we then ran a general linear model (GLM) style regression, in which the
257 neural RDM was modeled as a linear combination of two model RDMs: Category
258 (Animacy) and Visual (Fig. 4A), resulting in two beta-weights describing the individual
259 contribution of each model to the neural dissimilarity. Finally, the resulting beta-maps for
260 all participants were entered into a second-level analysis in SPM 12, producing a whole-
261 brain maps reflecting the contributions of Animacy and Visual information to the VTC
262 response patterns (Fig. 4B). Additional details about the Searchlight procedure can be
263 found in Proklova et al., 2016. We also ran two additional versions of the RSA using
264 smaller subsets of a full stimulus set. For example, in order to exclude animals with faces
265 from the analysis, we removed the entries of the neural and model RDMs that
266 corresponded to 6 animals with faces, resulting in smaller 12 x 12 RDMs that included
267 only faceless animals and inanimate objects. The same logic was applied when excluding
268 faceless animals from the analysis. Apart from this, the RSA procedure was identical to
269 the one described above.

270

271 *ROI-based RSA*

272 In the ROI-based RSA, the pairwise neural dissimilarity was measured in the same way
273 as in Searchlight RSA described above, but instead of doing it for each voxel in the brain,
274 it was done in each of the four ROIs (Fig. 3A). After constructing the neural RDMs, we
275 then correlated them with 5 model RDMs that corresponded to Animacy, Face presence,
276 Movement speed, Thoughtfulness, and Visual dissimilarity (Fig. 3B). The neural and
277 model RDMs were normalized before running this analysis. The correlation values were
278 t-tested against zero, and the resulting p-values were corrected for multiple comparisons
279 (Bonferroni correction, 3 analyses x 4 ROIs x 5 correlations = 60 tests, adjusted alpha
280 threshold $0.05/60 = 0.0008$). The Animacy RDM was defined by assigning 1 (maximum
281 dissimilarity) to pairs of stimuli belonging to the same category (animate or inanimate),
282 and 0 (minimum dissimilarity) to pairs from different categories. The Faceness, Thoughts
283 and Movement RDMs were defined based on the behavioral ratings from the ratings
284 experiment described above, correlating the ratings for each pair of images. Inanimate
285 objects (not included in the ratings experiment) were assigned a rating of zero. Finally,

286 the Visual RDM was constructed using the reaction times from the behavioral visual
287 search task described earlier. For each pair of stimuli, we used the inverse reaction time
288 (1/RT) for the corresponding target-distractor pair as a corresponding entry to the visual
289 RDM. Longer reaction times (indicating higher visual similarity) were thus reflecting lower
290 visual dissimilarity.

291

292 *Multidimensional scaling*

293 To visualize the relationship between the stimuli representations in each ROI, we
294 performed multidimensional scaling (MDS) using the *cmdscale* function in Matlab r2018b,
295 The MathWorks, Natick, MA.

296

297 **Results**

298 *Behavioral results*

299 We first wanted to check if our pre-selected faceless animals were indeed perceived as
300 faceless by participants. In our design, each of the six faceless animals (e.g., a worm)
301 was matched with a similarly shaped animal that had a more distinct face (e.g., a snake).
302 For each participant, we took the difference between the two “face presence” ratings for
303 each stimulus pair. The positive difference meant that an animal with a face received a
304 higher “face presence” rating than a corresponding faceless animal. As seen in the upper
305 panel of Figure 2D, this was the case for all six animal pairs. By contrast, there was no
306 significant difference in familiarity between animals with and without faces for each of the
307 six stimulus pairs (Fig. 2D, lower panel).

308 Next, we averaged the ratings across participants and arranged the animal stimuli
309 on a scale from the lowest to highest “face presence” rating (Fig. 2E), revealing that, on
310 average, all of the six preselected faceless animals were indeed rated lower on the
311 “faceness” scale than the six animals with faces ($p < 0.0001$). The same analysis was
312 performed with the ratings of movement speed and capacity for thought (see Fig. 2E). On
313 average, animals with faces were rated as faster moving and more capable of thought
314 compared to faceless animals ($p = 0.02$ and $p = 0.002$, respectively.) The ratings of head
315 and eye presence were almost identical to face presence ratings and were not further
316 analysed.

317 Perhaps unsurprisingly, familiarity ratings correlated positively with face presence
318 ($r = 0.42$), movement speed ($r = 0.64$), and thoughtfulness ($r = 0.54$). It can be challenging
319 to disentangle familiarity from these factors because we tend to be more familiar with
320 animals that move and think. As mentioned above, however, there was no significant
321 difference in familiarity between animals from the same shape pair with and without a
322 face.

323

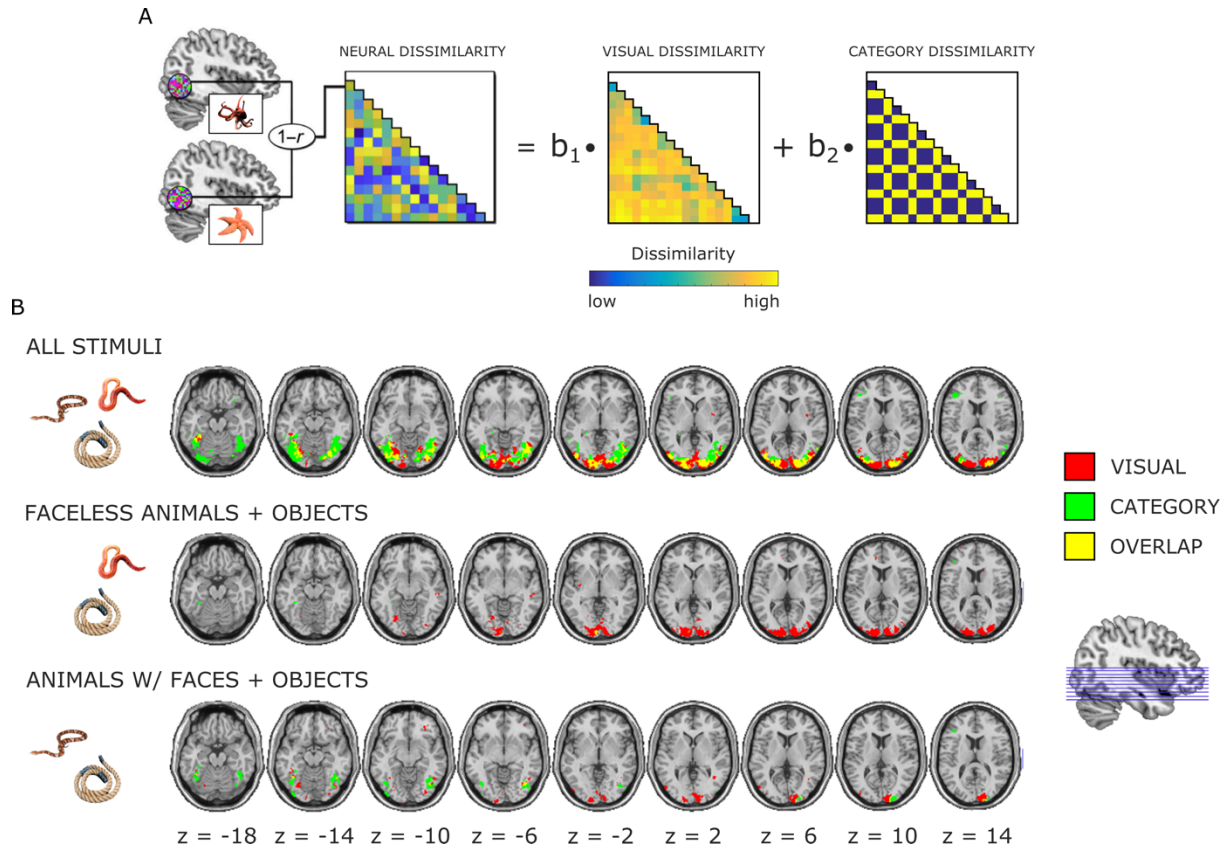
324 *RSA Searchlight Results*

325 Our next goal was to replicate the animate/inanimate distinction in VTC representations
326 found in earlier studies (e.g. Proklova et al, 2016) and to explore the possibility that these
327 results could have been driven by animal faces.

328 We performed the representational similarity searchlight analysis (identical to the
329 one used in Proklova et al, 2016) to reveal independent contributions of Animacy and
330 Visual models to object representations on the whole brain level (Fig. 4, see Methods for
331 the description of the analysis.) Importantly, this analysis was run three times: once with
332 all stimuli (Fig. 4B, upper row), once with only faceless animals and inanimate objects
333 (Fig. 4B, middle row), and once only with animals with faces and inanimate objects (Fig.
334 4B, bottom row). This allowed us to directly examine the contribution of animal faces to
335 the animate/inanimate distinction in the brain.

336 The first analysis, which included all the stimuli, revealed bilateral clusters in the
337 VTC in which Animacy model correlated significantly with the neural RDM (Fig. 4B, upper
338 row), with local peaks in left fusiform gyrus (19,096 mm³, peak coordinates: $x = -38$, $y =$
339 -62 , $z = -14$) and right fusiform gyrus (22,624 mm³, peak coordinates: $x = 40$, $y = -58$, z
340 $= -14$). However, when animals with faces were excluded from the analysis, we observed
341 only a small cluster in the left fusiform gyrus (168 mm³, peak coordinates: $x = -36$, $y = -$
342 44 , $z = -16$) in which the neural dissimilarity correlated significantly with category
343 (animacy) dissimilarity (Fig. 4B, middle row). As fewer stimuli (and fewer trials) were
344 included in this analysis, there is a possibility that the reduced animacy-related
345 information was observed due to reduced power. To control for this possibility, we then
346 ran the final searchlight analysis in which animals with faces were included in the analysis
347 and faceless animals excluded, which involved the same number of trials as the previous

348 analysis (Fig. 4B, bottom row). The results showed that the animacy information was
 349 again robust in the VTC in two clusters with peaks in left inferior temporal cortex (1,992
 350 mm³, peak coordinates: x = -48, y = -64, z = -10) and right fusiform gyrus (3,336 mm³,
 351 peak coordinates: x = 32, y = -72, z = -18). This suggests that the reduced animacy-
 352 related response observed in absence of animals with faces was not due to reduced
 353 power, but specifically to the absence of a face. Together, these results show that
 354 including animals with faces in the analysis leads to much more robust animate/inanimate
 355 distinction in the VTC. They also raise a possibility that the animacy information reported
 356 in previous studies (e.g., Proklova et al., 2016) could largely be an artefact of faces in the
 357 animate stimuli.
 358



359
 360 **Figure 4. GLM Searchlight.** **A**, Schematic of the analysis. For each voxel, we defined a
 361 100-voxel neighborhood around it and modeled the neural dissimilarity in this region as
 362 the linear combination of Visual and Category (Animacy) dissimilarity. **B**, Searchlight
 363 results. Results of whole-brain group-averaged analyses testing the value of each
 364 predictor versus zero. The results show independent contributions of Visual (in red) and
 365 Category (in green) predictors to neural data. This analysis was run three times: first,

366 including all the stimuli (upper row), next, after removing animals with faces (middle row),
367 and, finally, including only animals with faces and inanimate objects were included
368 (bottom row).

369

370

371 *ROI-based RSA Results*

372 The searchlight analysis showed that animal faces clearly play an important role in the
373 VTC representations. What is it about face presence that is driving this effect, and is it
374 different in different sub-regions of the VTC? Since face presence is associated with many
375 different factors (from visual features to perceived intelligence and similarity to humans)
376 we wanted to examine this further, including new behavioral models that captured
377 different aspects of the stimuli. We also focused on four regions of interest (ROIs) that
378 were defined prior to and independent from the searchlight analysis, including a large
379 VTC ROI, the fusiform face area (FFA), a region sensitive to animate/inanimate distinction
380 independently of visual features (Animacy ROI, see Methods for the details of how it was
381 defined), and early visual cortex (EVC) as a control region.

382 For each ROI, we correlated the neural representational dissimilarity matrix (RDM)
383 with five model RDMs characterizing Animacy, Faceness (face presence), Movement
384 Speed, Thoughts, and Visual information. The correlation values were then tested against
385 zero, and p-values were corrected for multiple comparisons (Bonferroni correction). The
386 results are shown in Figure 5. Analogously to the searchlight analysis, the ROI-based
387 RSA was run 3 times using different subsets of stimuli in order to directly examine the
388 effect of animal faces on object representations in those regions.

389 We first ran this analysis with the full stimulus set (Fig. 5, left column). As expected,
390 the Visual model, but not the high-level ones, correlated significantly with the early visual
391 cortex RDM. In the VTC, the Animacy model showed the highest correlation with the
392 neural RDM, followed by the Movement and Thoughts models for which the correlations
393 were also significant. Interestingly, the correlation with the Faceness RDM did not survive
394 correction for multiple comparisons. In the FFA, although none of the correlations
395 survived multiple comparisons correction, the highest correlation was with the Faceness
396 model. Finally, in the Animacy ROI, all high-level models (apart from Visual) correlated
397 significantly with the neural RDM.

398 Next, to see whether these results were driven by animals with faces, we re-ran
399 the RSA after excluding all animals with faces from the analysis (Fig. 5, middle column).
400 The results showed overall lower correlation values in all the ROIs, and, crucially,
401 animacy information was not significant after correcting for multiple comparisons in both
402 VTC and Animacy ROI. Interestingly, even in absence of faces, both Movement and
403 Thoughts RDMs still correlated significantly with the neural RDM in Animacy ROI.

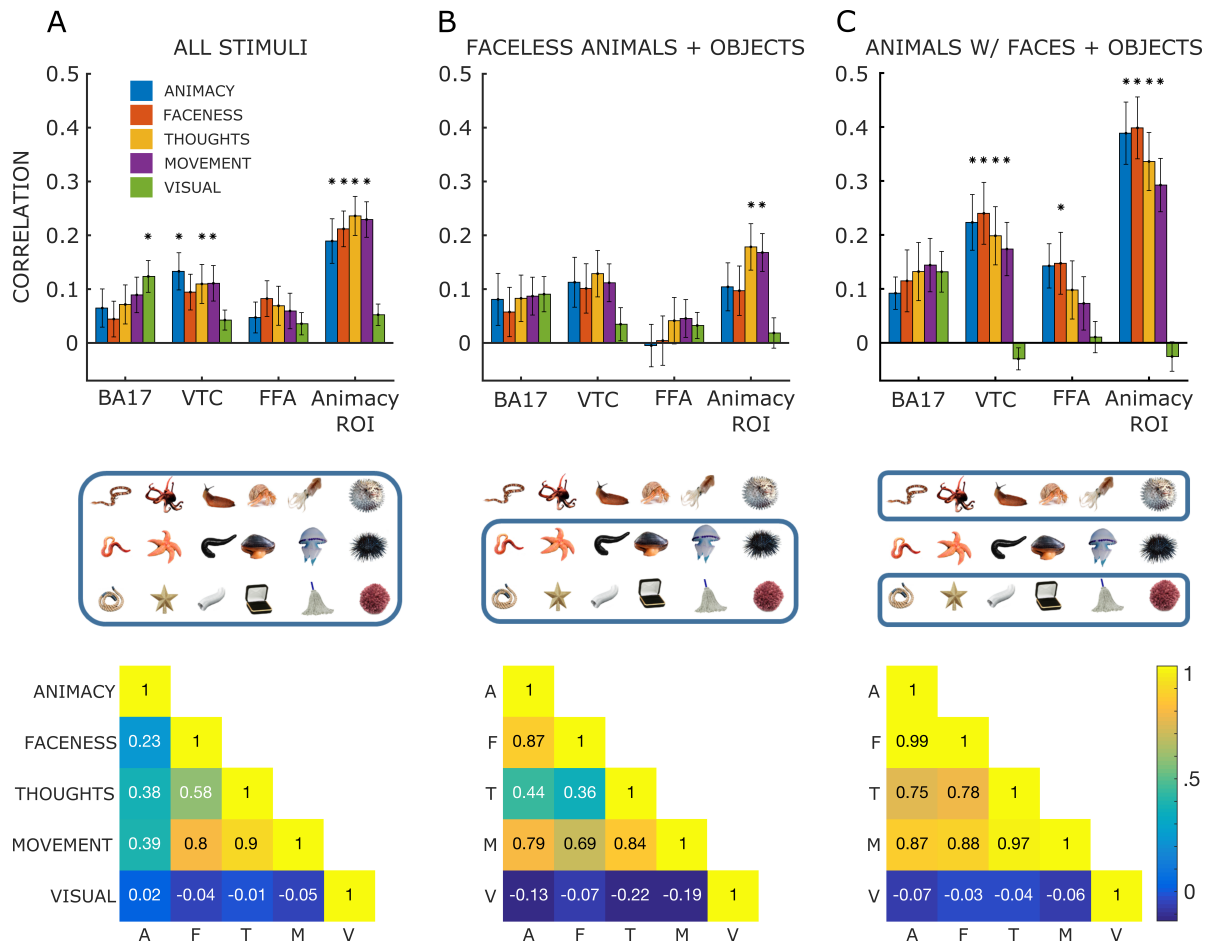
404 Finally, we re-ran the analysis after removing faceless animals and looking only at
405 animals with faces and inanimate objects (Fig. 5, right column). Note that Animacy and
406 Faceness RDMs were almost perfectly correlated in this condition. Strikingly, this led to
407 much higher correlations in VTC and the Animacy ROI with the Animacy RDM and all the
408 other high-level models.

409 Using different subsets of stimuli in the three analyses meant that the target RDMs
410 and correlations between them also changed. Pairwise correlations between target RDMs
411 are shown in the bottom row of Figure 5. In all three analyses, the Visual RDM did not
412 correlate highly with the remaining high-level models. The correlation between Animacy
413 and Faceness models was 0.23 for the full stimulus set, increased to 0.87 when animals
414 with faces were excluded (likely driven by the fact that in both models the inanimate
415 objects had a rating of zero), and was close to perfect (0.99) in the final analysis when
416 only animals with faces and inanimate objects were included.

417 At first glance, these results, in line with the Searchlight, seem to suggest that
418 faces heavily influence the representations of animacy in these regions: when animals
419 with faces are excluded, animacy information in VTC is not significant, and when they are
420 present, it is strongly pronounced. If this were the case, however, we would expect to see
421 higher correlations between the Faceness model and the VTC RDM, which was not the
422 case. Instead, this result seems to be driven by something other than faces (but
423 something that correlates with the presence of a face) – in this case, movement speed
424 and capacity to think. Indeed, even when animals with faces were excluded, the
425 Movement and Thoughts models (but not the binary Animacy model) correlated
426 significantly with the neural RDM in the Animacy ROI. This suggests that the
427 animate/inanimate distinction in this region is influenced not so much by the presence of

428 faces, but rather by other properties that correlate with face presence, such as the
 429 capacity to move and think.

430



431

432 **Figure 5. ROI RSA results.** Top row: RSA results in four ROIs including all the stimuli
 433 (A), the same analysis repeated after removing animals with faces from the analysis (B),
 434 and after animals with faces were included and faceless animals excluded from the
 435 analysis (C). Asterisks indicate significance after correcting for multiple comparisons.
 436 Middle row: stimuli that were included in each type of analysis (in blue frame). Bottom
 437 row: pairwise correlations between model representational dissimilarity matrices (RDMs)
 438 used in the three analyses.

439

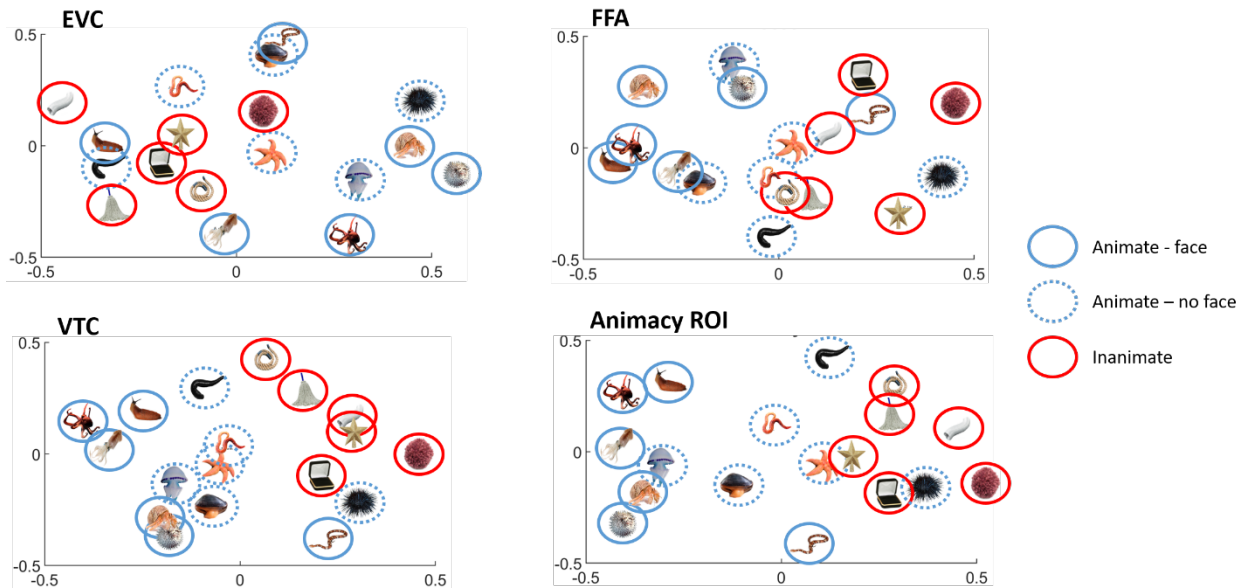
440 *Multidimensional scaling results*

441 Finally, we performed multidimensional scaling to visualize the representational structure
 442 in each ROI (Fig. 6). In this analysis, the images that are represented similarly in a given
 443 ROI end up closer together on a 2D plane. As expected, the representations in the EVC
 444 did not show clustering based on Animacy (compared to the other ROIs) and instead

445 seemed to have reflected visual properties: elongated stimuli, such as the tube, slug, and
446 leech, clustering together on the left and roundish objects clustered on the right. In both
447 the VTC and the Animacy ROI, however, the animate/inanimate distinction was
448 pronounced, with inanimate objects clustering together separately from animals (Fig. 6).
449 Intriguingly, this analysis revealed a kind of gradient in those particular ROIs: animals with
450 faces on one side, inanimate objects on the other side, and faceless animals in between.
451 This explains the results of the RSA, showing how including only animals with faces in
452 the analysis makes the distinction between animate and inanimate objects in those
453 regions more pronounced.

454

455



456

457 **Figure 6. Multidimensional scaling results.** Representational structure in the four ROIs
458 revealed by multidimensional scaling, showing how animals with faces (solid blue circles),
459 faceless animals (dashed blue circles), and inanimate objects (red circles) are
460 represented with respect to each other in each ROI. Stimuli that are represented similarly
461 in a given ROI are shown close to each other in 2-dimensional space.

462

463

464

465 **Discussion**

466 We investigated the contribution of animal faces to the animate-inanimate distinction that
467 has been revealed in ventral temporal cortex (VTC) across many studies (e.g.,
468 Kriegeskorte et al., 2008; Grill-Spector and Weiner, 2014; Bracci and Op de Beeck, 2016;
469 Proklova et al., 2016). Unlike previous studies, we systematically controlled for face
470 presence using images of real animals, half of which had a face and half of which did not,
471 as opposed to obscuring the face or using different viewpoints in which a face is turned
472 away (in which case our knowledge of how an animal usually looks could lead to filling-in
473 effects). The initial searchlight analysis revealed that when animals with faces were
474 removed from the analysis, the animate/inanimate distinction almost disappeared.
475 However, further ROI-based representational similarity analysis revealed that Movement
476 and Thoughts models significantly correlated with activity in a sub-region of the VTC –
477 even after faces were removed. Together, these results suggest that the
478 animate/inanimate distinction in the VTC is largely driven not by the presence of faces or
479 animacy *per se*, but rather by perceived agency (a combination of the ability to move and
480 the ability to think) that is correlated with these factors.

481

482 ***Visual vs. Conceptual Features***

483 How do our findings relate to the visual vs. conceptual features debate about the nature
484 of the animate/inanimate distinction in VTC object representations (Peelen & Downing,
485 2018; Bracci et al., 2019)? Although we did not address this question directly, our findings
486 do shed some light on the issue. Faces are both a visual and a conceptual feature. They
487 have characteristic visual features, such as eyes and mouth, and even simplistic smiley
488 faces and two dots above a line will elicit a response in the fusiform face area (Caldara
489 et al., 2006; Kim et al., 2016). Seeing faces in inanimate objects, such as clouds,
490 mountains, and tree trunks, is a common experience, underscoring the important
491 biological function of face recognition (Wardle et al., 2020). At the same time, faces
492 convey rich conceptual information, including similarity to humans (Sha et al., 2015;
493 Contini et al., 2019), emotion, and, in case of human faces, information about gender,
494 race, and age (Dobs et al, 2019). Faces are a powerful cue to whether something is
495 animate and, potentially, possessing agency and intelligence. Our findings suggest that
496 VTC processes not just the visual appearance of a face, but also higher-level information

497 for which faces are a proxy: specifically, animal's capacity for agency (movement and
498 thought). Moreover, we found that animal faces are not necessary for eliciting
499 animate/inanimate distinction in the VTC, in line with earlier studies (Chao et al, 1999;
500 Martin & Weisberg, 2003; Long et al., 2018).

501 It could still be the case that the animacy-related activity observed when animals
502 with faces were excluded from the analysis was driven by some remaining visual features,
503 such as curvature, symmetry, and visual texture, that differentiated the animals without
504 faces from the inanimate objects. That explanation is unlikely to be the whole story,
505 however, since we selected our images in a way that minimized shape and texture
506 differences, which was confirmed by low correlation between Visual and Animacy models.
507 Moreover, in a previous study (Proklova et al., 2016), we showed that visual features such
508 as overall shape and texture are not driving the animate/inanimate distinction in the VTC.
509 Of course, we still have to rely on visual information in order to recognize faceless animals
510 as animals, but the fact that we observed high-level information about movement speed
511 and capacity for thought in the Animacy ROI strongly suggests that our conceptual
512 knowledge about an animal also comes into play in this region. Many of us have learned
513 through the media and real-life experiences that starfish, sea urchins, and other creatures
514 that at first could appear as inanimate are in fact animals. In other words, semantic
515 associations between the image and previous knowledge is likely driving the observed
516 activation in the animacy-sensitive regions when images of animals without faces are
517 presented.

518

519 ***A Gradient vs. a Dichotomy***

520 More and more studies show that a simple animate/inanimate dichotomy is not the best
521 way to explain the VTC representational geometry (Bracci et al., 2018; Contini et al., 2020;
522 Connolly et al., 2012; Sha et al., 2016; Carlson et al., 2013). Our findings also suggest
523 that animacy in the VTC is not all-or-none, but graded: from animals that are perceived
524 as more mobile and intelligent to animals that are perceived as less capable of movement
525 and thought (and thus more similar to inanimate objects). Other recent findings have also
526 found that agency is an important organizing principle for the VTC object representations
527 (Thorat et al., 2019; Haxby et al., 2020).

528 It has also been proposed that this continuum is driven by similarity to humans
529 (Connolly et al., 2012; Sha et al., 2016; Ritchie et al., 2020), which could explain why
530 inanimate objects such as robots and toys are represented similarly to animals in VTC
531 (Bracci et al., 2019; Contini et al., 2020). Our study did not address this directly, since by
532 design all the animate stimuli were quite dissimilar to humans. Our findings do point,
533 however, to the importance of perceived agency (a combination of ability to move and
534 intelligence) for object representations in VTC. Animals (or animal-like objects) that are
535 perceived as possessing agency are indeed more similar to humans, compared to
536 animals that do not move and have simpler nervous systems. That said, our results
537 suggest that an animal does not have to share visual features with humans or to be
538 evolutionary “closer” to them in order to be represented distinctly from inanimate objects
539 in VTC. Capacity to move and intelligence are very behaviorally relevant features when it
540 comes to perception and recognition of animals, regardless of how similar an animal is to
541 a human. After all, a snake shares few physical features with humans, but it is important
542 to recognize it as animate in order to avoid danger – and movement (as well as the face
543 presence) is an important cue.

544

545 ***Static vs moving stimuli***

546 Like most studies that have explored the distinction between activity related to animate
547 and inanimate objects in the VTC, we presented our participants with static images. Had
548 we used video displays of animals vs. non-animals, then the presence of self-movement
549 or agency certainly would have been a powerful cue to animacy (Martin and Weisberg,
550 2003; Haxby et al., 2020). In other words, self-movement could be as potent a cue for
551 animacy as faces. The few studies that have used stimuli that move like animals have
552 found a characteristic animate/inanimate distinction in the VTC (Martin and Weisberg,
553 2003). Moreover, even though we used only static images in our study, the presence of
554 faces in some of the images could easily invoke the concept of movement (and other
555 features associated with animacy). Thus, as we have already discussed, this nexus of
556 associated animacy features could explain why images of animals with faces are
557 represented as more “animate” in the VTC compared to images of faceless animals, thus
558 eliciting a stronger animate/inanimate distinction.

559

560 ***Implications for Future Investigations***

561 Our study does not speak to how and where the associations between faces and other
562 aspects of animacy are encoded. There is a possibility that high-level aspects of animacy
563 (e.g., agency) are first processed outside of VTC, and this information is then conveyed
564 back to the VTC via re-entrant pathways. Electrophysiological techniques such as M/EEG
565 could shed light on the time course of this process (Cichy & Oliva, 2020). Our findings
566 also suggest that any attempt to disentangle the factors contributing to the animate-
567 inanimate distinction (or gradient) in VTC should pay close attention to the animals that
568 are used as stimuli. It is critical to include a wide range of animate objects, not just more
569 typical, human-looking mammals. Moreover, this applies, not just to animals, but to any
570 object category that is being investigated. The use of large, diverse, and naturalistic
571 stimuli sets (Hebart et al., 2019, Nastase et al., 2020) is one way forward.

572 **(1331 words)**

573

574 **References**

- 575 Bracci, S., & Op de Beeck, H. (2016). Dissociations and Associations between Shape
576 and Category Representations in the Two Visual Pathways. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 36(2), 432–444.
- 578 Bracci, S., Ritchie, J. B., & Op de Beeck, H. (2017). On the partnership between neural
579 representations of object categories and visual features in the ventral visual pathway.
580 *Neuropsychologia*, 105, 153–164.
- 581 Bracci, S., Ritchie, J. B., Kalfas, I., & Op de Beeck, H. (2019). The ventral visual pathway
582 represents animal appearance over animacy, unlike human behavior and deep neural
583 networks. *Journal of Neuroscience*.
- 584 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
- 585 Caldara, R., Seghier, M. L., Rossion, B., Lazeyras, F., Michel, C., & Hauert, C. A. (2006).
586 The fusiform face area is tuned for curvilinear patterns with more high- contrasted
587 elements in the upper part. *Neuroimage*, 31(1), 313-319.
- 588 Caramazza, A., & Shelton, J. R. (1998). Domain-Specific Knowledge Systems in the
589 Brain: The Animate-Inanimate Distinction. *Journal of Cognitive Neuroscience*, 10(1), 1–
590 34.
- 591 Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only
592 for faces? *Neuroreport*, 10(14), 2945–2950.
- 593 Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain
594 Responses in Space and Time. *Neuron*, 107(5), 772–781.
- 595 Coggan, D. D., Baker, D. H., & Andrews, T. J. (2016). The Role of Visual and Semantic
596 Properties in the Emergence of Category-Specific Patterns of Neural Response in the
597 Human Brain. *ENeuro*, 3(4).
- 598 Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., Abdi,
599 H., & Haxby, J. V. (2012). The representation of biological classes in the human brain.
600 *Journal of Neuroscience*, 32(8), 2608–2618.
- 601 Contini, E. W., Goddard, E., Grootswagers, T., Williams, M., & Carlson, T. (2019). A
602 humanness dimension to visual object coding in the brain. *BioRxiv*, 648998.
- 603 Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over
604 time. *Nature Communications*, 10(1), 1–10.

- 605 Gobbini, M. I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., Guazzelli, M.,
606 & Pietrini, P. (2011). Distinct neural systems involved in agency and animacy detection.
607 *Journal of Cognitive Neuroscience*, 23(8), 1911–1920.
- 608 Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral
609 temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15(8), 536–
610 548.
- 611 Grill-Spector, K., Weiner, K. S., Gomez, J., Stigliani, A., & Natu, V. S. (2018). The
612 functional neuroanatomy of face perception: From brain measurements to deep neural
613 networks. *Interface Focus*, 8(4), 20180013.
- 614 Haxby, J.V., Gobbini, M.I., & Nastase, S.A. (2020) Naturalistic stimuli reveal a dominant
615 role for agentic action in visual representation. *Neuroimage*, 216:116561.
- 616 Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Coriveau, A., Van Wicklin, C., &
617 Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000
618 naturalistic object images. *PloS one*, 14(10), e0223792.
- 619 Kim, K. W., Lee, S. W., Choi, J., Kim, T. M., & Jeong, B. (2016). Neural correlates of
620 text-based emoticons: a preliminary fMRI study. *Brain and behavior*, 6(8), e00500.
- 621 Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., &
622 Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior
623 Temporal Cortex of Man and Monkey. *Neuron*, 60(6), 1126–1141.
- 624 Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level
625 categorical organization of the ventral stream. *Proceedings of the National Academy of
626 Sciences of the United States of America*, 115(38), E9015–E9024.
- 627 Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and
628 mechanical concepts. *Cognitive Neuropsychology*, 20:3-6, 575-587.
- 629 Maldjian J.A., Laurienti P.J., Kraft R.A., & Burdette J.H. (2003). An automated method for
630 neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets.
631 *Neuroimage*, 19, 1233–1239.
- 632 Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual
633 categorization. *Journal of Vision*, 12(11), 19–19.
- 634 Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of
635 experimental control in cognitive neuroscience. *PsyArXiv*. doi:10.31234/osf.io/whn6d.

636 Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal
637 Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Front.*
638 *Neuroinform.* 10:27.

639 Peelen, M. V., & Downing, P. E. (2017). Category selectivity in human visual cortex:
640 Beyond visual object recognition. *Neuropsychologia*, 105, 177–183.

641 Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object
642 shape and object category in human visual cortex: The animate–inanimate distinction.
643 *Journal of Cognitive Neuroscience*, 28(5), 680–692.

644 Ritchie, J. B., Zeman, A. A., Bosmans, J., Sun, S., Verhaegen, K., & Op de Beeck, H. P.
645 (2020). Untangling the animacy organization of occipitotemporal cortex. *BioRxiv*,
646 2020.07.17.206896.

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

650 Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization
651 in human ventral temporal cortex. *Elife*, 8.

652 Wardle, S. G., Taubert, J., Teichmann, L., & Baker, C. I. (2020). Rapid and dynamic
653 processing of face pareidolia in the human brain. *Nature Communications*, 11(1), 4518.

654

655

656

657