

1 **The representational dynamics of visual objects in rapid serial visual processing streams**

2

3 Tijl Grootswagers^{*a,b,c,1}, Amanda K. Robinson^{*a,b,c,2}, Thomas A. Carlson^{a,b,3}

4 *equal contribution

5

6 ^a School of Psychology, University of Sydney, NSW, 2006, Australia

7 ^b ARC Centre of Excellence in Cognition and Its Disorders, NSW, 2109, Australia

8 ^c Department of Cognitive Science, Macquarie University, NSW, 2109, Australia

9

10 ¹tijl.grootswagers@sydney.edu.au ²amanda.robinson@sydney.edu.au ³thomas.carlson@sydney.edu.au

11

12 **Abstract**

13 In our daily lives, we are bombarded with a stream of rapidly changing visual input. Humans have the
14 remarkable capacity to detect and identify objects in fast-changing scenes. Yet, when studying brain
15 representations, stimuli are generally presented in isolation. Here, we studied the dynamics of human
16 vision using a combination of fast stimulus presentation rates, electroencephalography and multivariate
17 decoding analyses. Using a presentation rate of 5 images per second, we obtained the representational
18 structure of a large number of stimuli, and showed the emerging abstract categorical organisation of
19 this structure. Furthermore, we could separate the temporal dynamics of perceptual processing from
20 higher-level target selection effects. In a second experiment, we used the same paradigm at 20Hz to
21 show that shorter image presentation limits the categorical abstraction of object representations. Our
22 results show that applying multivariate pattern analysis to every image in rapid serial visual processing
23 streams has unprecedented potential for studying the temporal dynamics of the structure of
24 representations in the human visual system.

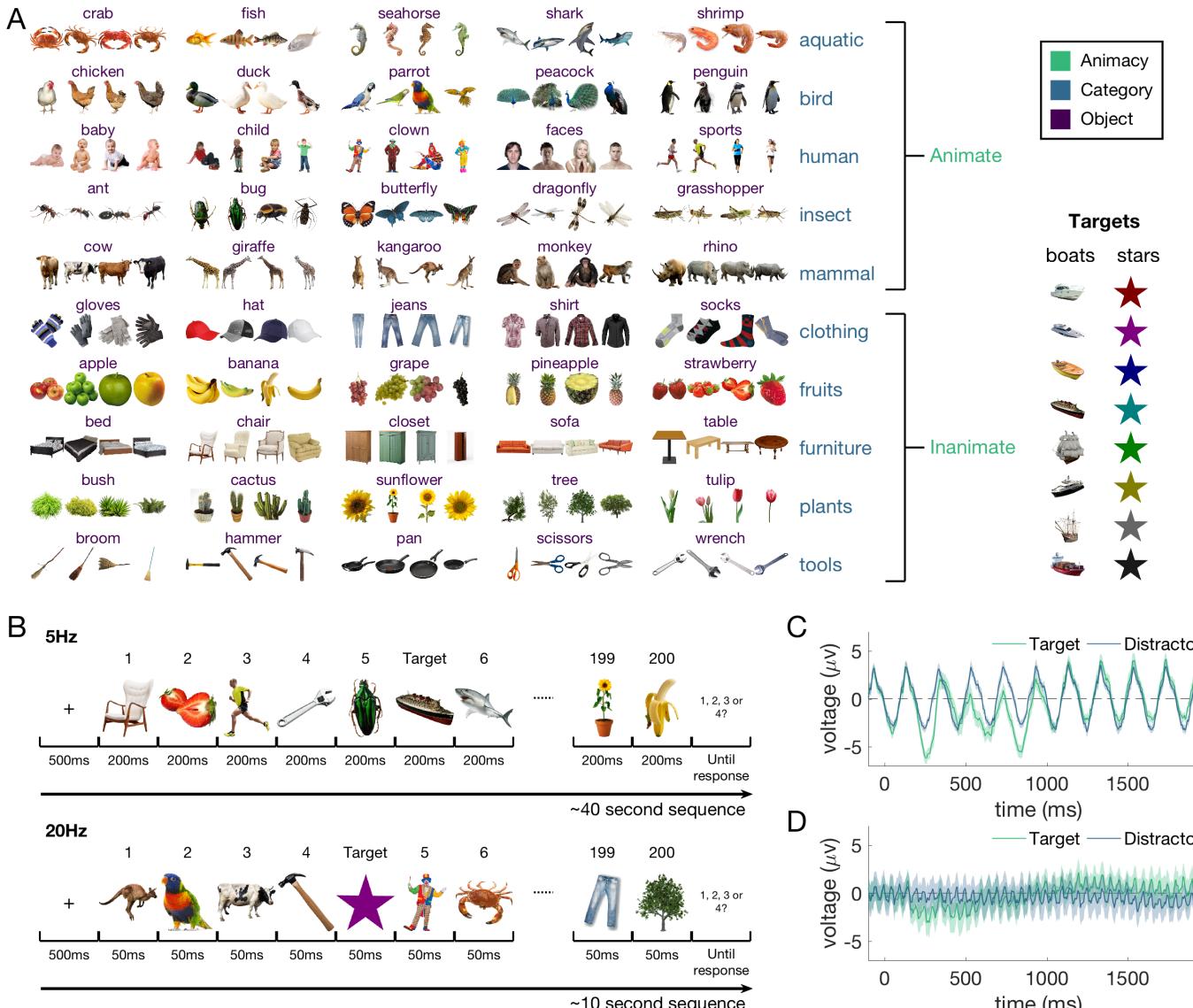
25 **Introduction**

26 The human brain can effortlessly extract abstract meaning, such as categorical object information, from
27 a visual image, and can do so in less than 200 milliseconds (Carlson, Tovar, Alink, & Kriegeskorte, 2013;
28 Cichy, Pantazis, & Oliva, 2014; Contini, Wardle, & Carlson, 2017; Keysers, Xiao, Földiák, & Perrett, 2001;
29 Mack, Gauthier, Sadr, & Palmeri, 2008; Mack & Palmeri, 2011; Potter, 1975, 1976; Potter, Wyble,
30 Hagmann, & McCourt, 2014; VanRullen & Thorpe, 2001). The temporal dynamics of the emerging
31 representation of visual objects has been studied extensively using multivariate decoding methods and
32 neuroimaging methods with high temporal resolution, such as EEG and MEG. In these experiments,
33 stimuli are generally presented with a large inter-stimulus interval (ISI) to avoid contamination from
34 temporally adjacent stimuli, typically around one second (Carlson et al., 2013; Cichy et al., 2014;
35 Grootswagers, Ritchie, Wardle, Heathcote, & Carlson, 2017; Isik, Meyers, Leibo, & Poggio, 2014;
36 Kaneshiro, Guimaraes, Kim, Norcia, & Suppes, 2015). This design allows the brain to process each
37 stimulus and avoids temporally overlapping stimulus representations. While such designs have yielded
38 important insights into the representational dynamics of object processing, in the natural world, we are
39 bombarded with a constant stream of changing visual input. The standard paradigm, in which stimuli are
40 presented in isolation with a large ISI, thus may not yield the most accurate description the temporal
41 dynamics of emerging object representations in the real world. One major advantage of multivariate
42 decoding methods (Grootswagers, Wardle, & Carlson, 2017; Haynes, 2015) is that they allow testing for
43 statistical dependencies in data without a resting baseline. Exploring representational dynamics using
44 decoding and fast visual presentation rates therefore offers unique potential for investigating visual
45 processing.

46

47 Here, we diverge from the traditional approach and propose a new method for studying the
48 representational dynamics of human vision. It has been shown previously that stimuli presented at high
49 presentation rates are all processed to some degree by the visual system and that their neural

50 representations can co-exist in the visual system (Marti & Dehaene, 2017; Mohsenzadeh, Qin, Cichy, &
51 Pantazis, 2018; Rossion, Torfs, Jacques, & Liu-Shuang, 2015; Rousselet, Fabre-Thorpe, & Thorpe, 2002).
52 Behavioural work has additionally shown that the human visual system can extract abstract information
53 from a visual stimulus at very fast presentation rates (Crouzet, Kirchner, & Thorpe, 2010; Keysers et al.,
54 2001; Macé, Thorpe, & Fabre-Thorpe, 2005; Mack et al., 2008; Mack & Palmeri, 2015; Marti & Dehaene,
55 2017; Potter, 1975, 1976; Potter et al., 2014; Rossion et al., 2015; Thorpe, Fize, & Marlot, 1996). In the
56 current study, we draw on this human capacity and study visual object recognition using fast stimulus
57 presentation rates and multivariate decoding analyses of EEG evoked responses (Grootswagers, Wardle,
58 et al., 2017). We used a rapid serial visual presentation (RSVP) paradigm to study the representations of
59 a large set of 200 visual objects presented at a speed of 5 images per second (5Hz; 200ms per image).
60 The objects were carefully selected to allow categorisation at three different levels of abstraction. The
61 high presentation rate enabled us to obtain 40 repetitions of 200 different stimuli in a short EEG session.
62 The increased power elicited by the faster image presentation rates allowed us to use a much larger
63 stimulus set than previous studies, and to analyse neural responses to all distractors, rather than a
64 single target, in the stream. We additionally examined the effect of higher level cognitive processes on
65 the emerging representations by having participants detect targets that were identifiable based on low-
66 level visual features or abstract categories in separate trials. In doing so, we could disentangle the
67 temporal dynamics of visual processing and categorical abstraction of non-target stimuli from target
68 selection processes. We successfully decoded different categorical contrasts for the 200 objects,
69 suggesting that individual stimuli were processed up to abstract categorical representations. Strikingly,
70 we found similar results in a follow-up experimental session, where we used a much higher presentation
71 rate of 20 images per second (20Hz; 50ms per image). The unprecedented ability to test such large
72 numbers of different stimuli in relatively short EEG scanning sessions shows great potential for studying
73 the dynamics of the structure of information in the human visual system.



75

76 Figure 1. Stimuli and design. A) Experimental stimuli. There were 200 images of objects (obtained from
 77 www.pngimg.com), organised in categories at three different levels: Animacy (animate, inanimate),
 78 category (10 categories e.g., mammal, tool, flower) and object (50 categories e.g., cow, dog, giraffe). In
 79 the experiment, participants were asked to count the number of target objects from two categories:
 80 boats and geometric star shapes, each with eight images. B) Experimental design. Trials consisted of all
 81 200 images presented in random order, with 1-4 targets interspersed throughout. Images were
 82 presented in 5Hz sequences (200ms each) in session 1, and 20Hz sequences (50ms each) in session 2.
 83 C,D) Subject-averaged event-related potentials (ERPs) at channel Oz for target and non-target images in
 84 the 5Hz (C) and 20Hz (D) sequences (shaded areas show the standard-error across subjects).

85

86

87

88

89 **Methods**

90 **Stimuli**

91 We collected a stimulus set of 200 visual objects from different categories. Stimuli were obtained from
92 the free image hosting website www.pngimg.com. The categories were manually selected, guided by
93 categorical hierarchies described in the literature (Caramazza & Mahon, 2003; Caramazza & Shelton,
94 1998; Carlson et al., 2013; Connolly et al., 2012; Grill-Spector & Weiner, 2014; Kiani, Esteky, Mirpour, &
95 Tanaka, 2007; Kriegeskorte, Mur, Ruff, et al., 2008; Mahon & Caramazza, 2011; Peelen & Caramazza,
96 2012; E. H. Rosch, 1973). There were two high level categories (animate, inanimate) consisting of 10
97 categories (5 animate, and 5 inanimate categories). Each of these 10 categories (e.g., mammal, tool,
98 flower) was further separated into 5 object categories (e.g., cow, dog, giraffe, etc.), which consisted of 4
99 images each (Figure 1a). During the experiment, participants were instructed to count target stimuli
100 (Figure 1b). To examine how attending to different features of the stimuli affected the emerging
101 representations, we used two different sets of target stimuli. The target stimuli were either boats, or
102 geometric star shapes, and there were eight exemplars of each target type (Figure 1 – inset). We
103 hypothesized that detecting the star shapes among the other objects was possible using low level visual
104 cues, while for recognising boat targets, it was necessary to process stimuli to a more abstract
105 categorical level.

106

107 **Participants and experimental procedure**

108 Participants were 16 adults recruited from the University of Sydney (5 females; age range 18-38 years)
109 in return for payment or course credit. The study was approved by the University of Sydney ethics
110 committee and informed consent was obtained from all participants. Participants viewed 40 sequences
111 of objects, each lasting between 40.2 - 40.8 seconds (depending on the number of targets in the
112 sequence). In each sequence, the 200 stimuli were presented in random order, for a duration of 200ms
113 each with no gap between successive images (5Hz). In addition to the 200 stimuli, target stimuli were

114 inserted throughout the sequence (Figure 1b). In half of the sequences, the target stimuli were boats,
115 and in the other sequences, the target stimuli were geometric stars (Figure 1). A random number
116 between 1 and 4 targets were presented in the sequence, with the condition that targets could not
117 appear within the first 10 or last 10 images, and ensuring there were at least 12 non-target stimuli
118 between subsequent targets. At the start of each sequence, participants were prompted to count the
119 number of targets in the sequence (“Count the boats in the trial” or “Count the stars in the trial” in
120 random order) and the 8 potential targets were shown. They were instructed to respond at the end of
121 the sequence using a 4-way button box. After each sequence, participants received feedback. They
122 started the next sequence with a button press. This session lasted approximately 40 minutes in total.
123 After a short break, the second experimental session started, and participants performed another 40
124 sequences using the same procedure as session one, except that the images were presented for only
125 50ms (a presentation speed of 20Hz). The second session lasted about 10 minutes.

126

127 **EEG recordings and preprocessing**

128 Continuous EEG data were recorded using a BrainVision ActiChamp system, digitized at a 1000-Hz
129 sample rate. The 64 electrodes were arranged according to the international standard 10–10 system for
130 electrode placement (Oostenveld & Praamstra, 2001). During recording, all scalp electrodes were
131 referenced to Cz. Preprocessing was performed offline using EEGLab (Delorme & Makeig, 2004). Data
132 were filtered using a Hamming windowed FIR filter with 0.1 Hz highpass and 100Hz lowpass filters, and
133 were downsampled to 250Hz. No further preprocessing steps were applied, and the channel voltages at
134 each time point were used for the remainder of the analysis. Epochs were created for each stimulus
135 presentation (except targets) ranging from [-100 to 1000ms] relative to stimulus onset. We initially had
136 used the same range for target-distractor decoding but found that this window did not capture the full
137 process. Therefore, for comparing targets versus distractors, we created larger epochs ranging from [-
138 100 to 2000ms] relative to the onset of a target. For each target t, we selected at random another

139 distractor in the same sequence and created a matching epoch relative to the onset of that distractor.
140 Choosing distractors in this way meant that the number of targets and distractors were balanced and
141 matched per sequence (and chance level accuracy is 50%) and that the neural representations of targets
142 and distractors were unlikely to overlap in a consistent manner. Event-related potentials (Figure 1C&D)
143 for both the targets and non-targets exhibited clear signal at the presentation frequencies (see Figure S1
144 for the associated scalp maps and amplitude spectra).

145

146 **Decoding analysis**

147 We applied an MVPA decoding pipeline (Grootswagers, Wardle, et al., 2017; Oosterhof, Connolly, &
148 Haxby, 2016) to the EEG channel voltages, consisting of a regularised linear discriminant analysis (LDA)
149 classifier applied in an exemplar-by-sequence-cross-validation approach. Decoding was performed
150 within subject, and the results were analysed at the group level. This pipeline was applied to each
151 stimulus presentation epoch in the sequence to investigate object representations in fast sequences. To
152 investigate the temporal dynamics of target selection, we compared neural responses to targets with
153 those to non-target distractor stimuli. Classifiers were then trained to distinguish targets from non-
154 targets separately for the 5Hz and 20Hz sequences, and for boat and star target sequences.

155

156 We investigated object representations for the 200 non-target images using multiple categorical
157 distinctions. First, we decoded three contrasts that impose different amounts of categorical abstraction.
158 At the highest level, we decoded animacy (i.e., animate versus inanimate objects). The next contrast was
159 the category tier (10 classes, e.g., mammal, insect, furniture, tool, etc.) where we decoded all 45
160 possible pairwise combinations. The lowest categorical level was the object level (50 classes, e.g., cow,
161 butterfly, table, hammer, etc.). Here, we decoded all 1225 possible pairwise object combinations (i.e.,
162 cow versus butterfly, cow versus table, etc.). Finally, at the lowest level, we investigated image-level
163 representations by decoding all 19900 possible pairwise combinations of the 200 stimuli. We report the

164 mean pairwise classification accuracies, so that chance-level accuracy for all comparisons is at 50%,
165 which aids comparing accuracies across contrasts.

166

167 To investigate similarities in underlying object representation signals between the 5Hz and 20Hz
168 presentations, we used a temporal generalisation approach (Carlson, Hogendoorn, Kanai, Mesik, &
169 Turret, 2011; King & Dehaene, 2014; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008). To test
170 generalisation between the conditions, we trained classifiers on all time points in the data from the 5Hz
171 sequences and tested them on all time points in the data from the 20Hz sequences. We repeated this
172 for the inverse (training on 20Hz and testing on 5Hz), and averaged the resulting time-generalisation
173 matrices (Kaiser, Azzalini, & Peelen, 2016).

174

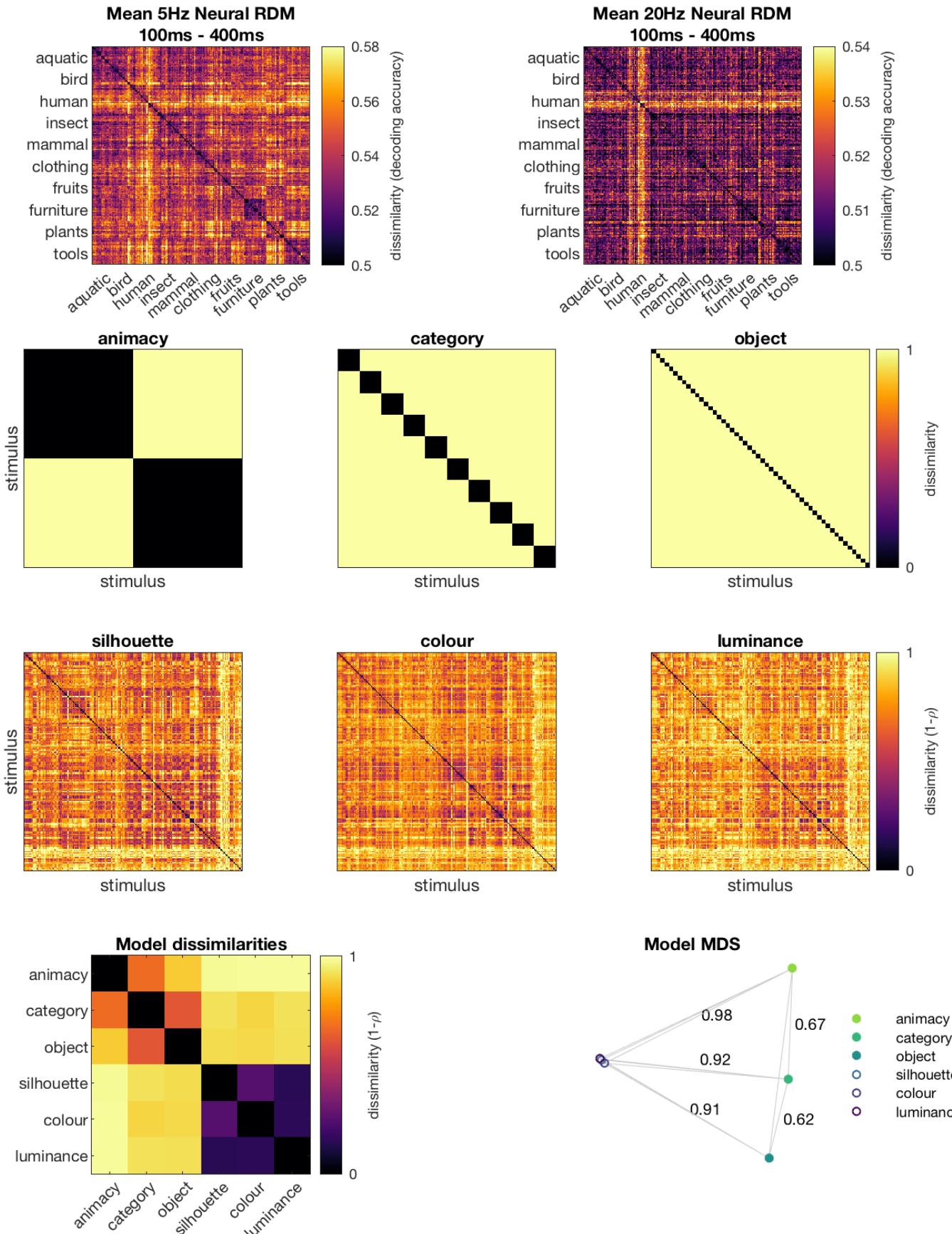
175 All steps in the decoding analysis were implemented in CoSMoMVPA (Oosterhof et al., 2016). For the
176 categorical contrasts that grouped more than one image, we used an image-by-sequence-cross-
177 validation scheme so that identical images were not part of both training and test set (Carlson et al.,
178 2013; Grootswagers, Wardle, et al., 2017). This was implemented by first splitting the data into four
179 sets, where the first set consisted of the first images from each of the 50 object categories (i.e., cow-1,
180 table-1 etc.), the second set of the second images (i.e., cow-2, table-2 etc.), etc. One of these sets was
181 used as test data, and the other three as training data for the leave-one-sequence out cross-validation,
182 where all data from one sequence was used as test data, and data from the remaining sequences as
183 training data. For each decoding contrast, this resulted in 160 (4 images by 40 sequences) cross-
184 validation partitions. Where image-by-sequence cross-validation was not possible (i.e., image-level and
185 target-distractor decoding), we used a leave-one-sequence-out cross-validation scheme, where all
186 epochs from one sequence were used as test set, resulting in 40 cross-validation partitions. We used a
187 linear discriminant analysis (LDA) classifier (implemented in CoSMoMVPA) and report the mean cross-
188 validated decoding accuracy.

189

190 **Representational Similarity Analysis**

191 To study the emerging representational structure of our 200 stimuli, we analysed our data using the
192 Representational Similarity Analysis (RSA) framework (Kriegeskorte & Kievit, 2013; Kriegeskorte, Mur, &
193 Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008), which allows comparing models of object
194 representations. The decoding results at the image level were organised into a 200 by 200 neural
195 representational dissimilarity matrix (RDM), which for each pair of images, contained the mean cross-
196 validated decoding accuracy (images that evoke more dissimilar neural responses are better decodable).
197 One neural RDM was created for each subject, and each time point (group mean RDM at 100-400ms
198 shown in Figure 2, top row). We compared the neural RDMs to six candidate models; first, we created
199 one model for each of the three categorical levels, grouping images from the same category (Figure 2,
200 second row). We also used three low-level image feature control models (Figure 2, third row), which
201 were created by correlating the vectorized experimental images. The models consisted of an image
202 silhouette similarity model, which is based on the binary alpha layer of the stimuli and is a good
203 predictor of differences in brain responses (Carlson et al., 2011; Teichmann, Grootswagers, Carlson, &
204 Rich, 2018; Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016)), a model based on
205 the CIELAB-colour values of the stimuli, and a model based on the difference in luminance of the stimuli.
206 Figure 2 shows the candidate models and the correlation distance between each of the candidate
207 models (bottom row). The small correlations between the categorical models and the low-level feature
208 models suggests that there was little overlap between the low-level features and categorical
209 organisations in the stimulus set. To quantify the unique contributions of all models to the neural
210 dissimilarities, we modelled the time-varying neural RDMs of each subject as a linear combination of the
211 candidate models using a GLM (Oosterhof et al., 2016; Proklova, Kaiser, & Peelen, 2017); for each time
212 point, the lower triangles of the neural RDM and candidate models were vectorised, and regression
213 coefficients were obtained for all candidate models. This resulted in one beta estimate for each model,

214 subject, and time point. We then analysed at the group level the mean beta estimates across subjects.
215 To visualise the dynamic representational structure, at each point in time, we created a two-
216 dimensional embedding of all 200 images. To compute the two-dimensional embedding, we applied t-
217 SNE (Maaten & Hinton, 2008) to the mean neural RDMs. This approach finds an embedding of the multi-
218 dimensional space in a two-dimensional representation so that the distances between points reflect
219 their multidimensional pattern dissimilarities as best as possible.



220

221 Figure 2. Candidate models used in the RSA. Top row: time-averaged neural RDMs for the 5Hz and 20Hz
 222 conditions. Each point in the 200 by 200 matrix represents the dissimilarity (here: decoding accuracy)
 223 between a pair of images. Second row: categorical models predict that responses to stimuli from the

224 same category are more similar than responses to stimuli for different categories. Third row: image
225 properties entered the regression as control models to quantify the contribution of low-level visual
226 differences to the neural dissimilarities. Bottom row left: dissimilarities (1-correlation) between all
227 candidate models. The order of the images in the 200 x 200 RDMs are the same as in Figure 1. Bottom
228 row right: model dissimilarities projected in a 2-dimensional space using classical multi-dimensional
229 scaling, which returns a configuration so that the distance between points approximates their
230 dissimilarities. Annotated are the dissimilarities (1-correlation) between category-animacy and category-
231 object, and between the silhouette model and all three categorical models.
232

233 **Statistical inference**

234 In this study, we used Bayes factors (Dienes, 2011; Jeffreys, 1998; Rouder, Speckman, Sun, Morey, &
235 Iverson, 2009; Wagenmakers, 2007) to determine the evidence for the null and alternative hypotheses.
236 For the alternative hypothesis of above-chance decoding or correlation, a uniform prior was used
237 ranging from the maximum value observed during the baseline (before stimulus onset) up to 1 (e.g.,
238 100% decoding). For testing a non-zero difference between decoding accuracies, a uniform prior was
239 used ranging from the maximum absolute difference observed during the baseline up to 50% (0.5). We
240 then calculated the Bayes factor (BF) which is the probability of the data under the alternative
241 hypothesis relative to the null hypothesis. We thresholded BF>3 and BF>10 as substantial and strong
242 evidence for the alternative hypothesis, and BF<1/3 and BF<1/10 for substantial/strong evidence in
243 favour of the null hypothesis (Jeffreys, 1998; Wetzels et al., 2011). BF that lie between 1/3 and 3
244 indicate insufficient evidence for either hypothesis.
245

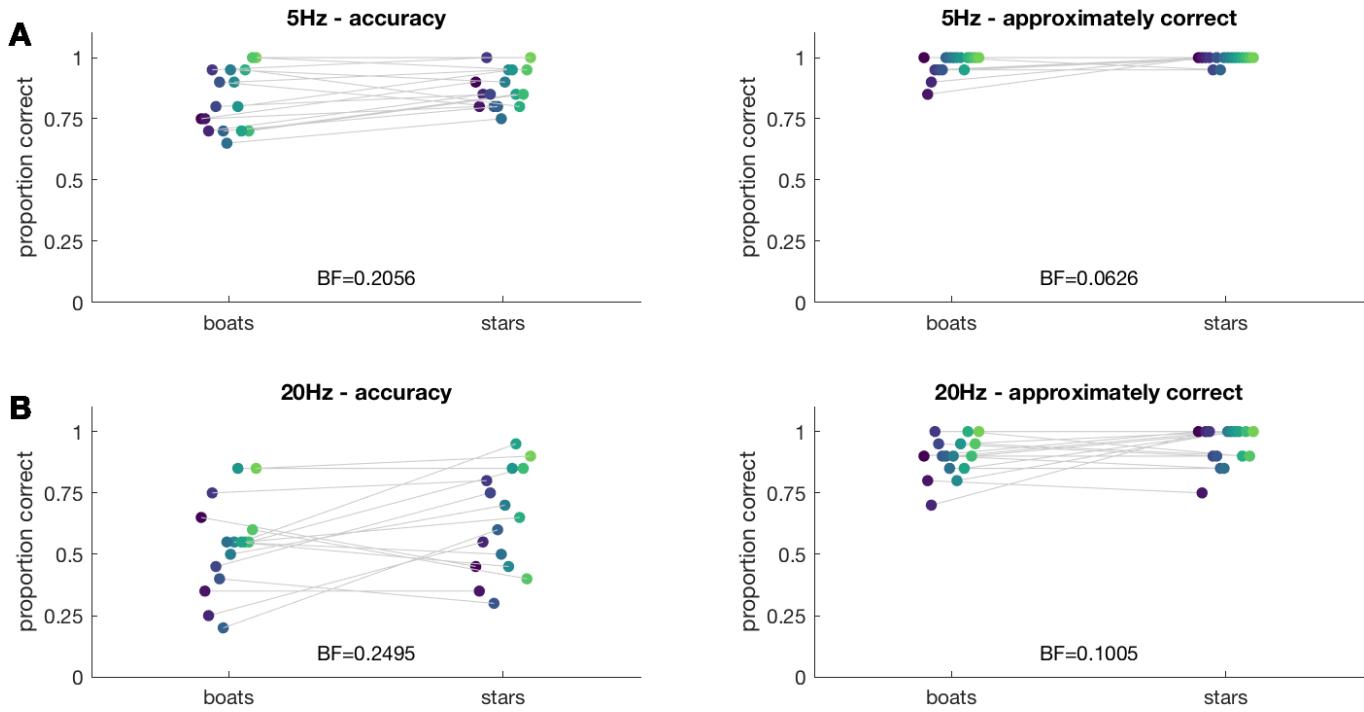
246 **Results**

247 We examined the representational dynamics of 200 different visual objects (Figure 1A), presented in
248 5Hz and 20Hz sequences (Figure 1B) using EEG. During the sequences, participants detected targets
249 (boats or stars).

250

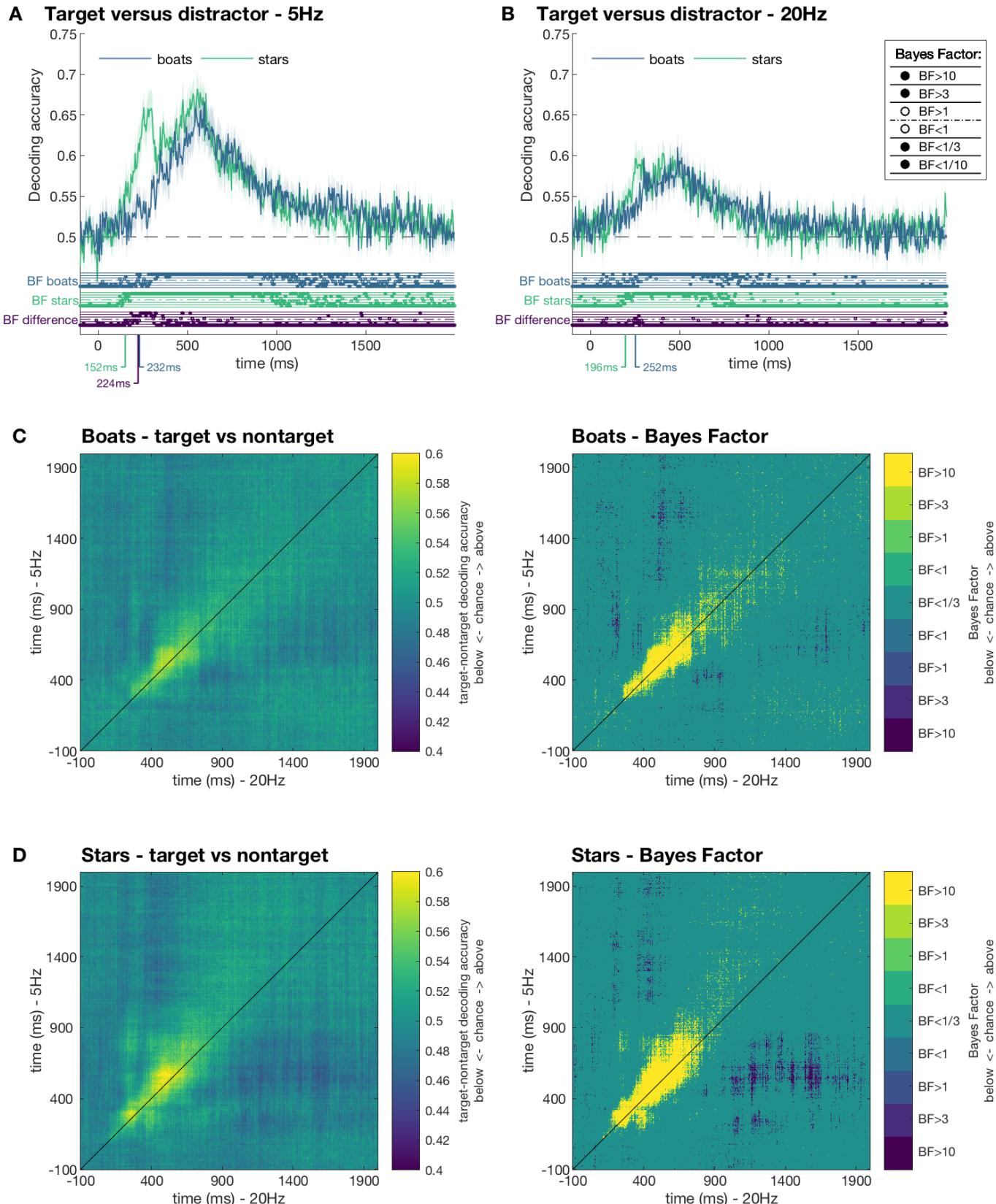
251 **The effect of target type and target selection**

252 Participants were generally above chance (25%) at detecting targets (boats or stars) in the 5Hz and 20Hz
253 sequences (Figure 3A-B). There was no difference in performance between the boat and star conditions
254 (all $BF < 1/3$). On incorrect trials, responses often differed no more than one from the correct answer
255 (Figure 3, right columns). This indicates that in general, participants missed at most one target when
256 they responded incorrectly.



257
258 Figure 3: Behavioural results of target detection performance for 5Hz (A) and 20Hz (B) experimental
259 sessions. Left columns show the mean proportion of correct responses for each participant separately
260 for boat target sequences and star target sequences. Right columns show the mean approximately-
261 correct (i.e., response differed by at most 1 from the correct answer) accuracy for each participant.
262 Bayes Factors (BF) comparing mean accuracies between the boat and star sequences are listed above
263 the x-axis.
264

265



266

267 Figure 4. Decoding target versus distractor. For each target, a distractor was randomly selected from the
 268 same sequence, and classifiers were trained on target versus distractor. Plots show the mean leave-one-
 269 sequence-out cross-validated accuracy for the 5Hz condition (A), and the 20Hz condition (B). Shaded
 270 areas show the standard error of the mean across participants. Results are shown separately for boat
 271 target sequences and star target sequences. Dots below plots indicate thresholded Bayes Factors (BF,

272 see inset) for the boat (top row) and star (middle row) sequences compared to chance and for the
273 difference between boat and star sequences (bottom row). Annotated below the x-axis are the time
274 points where the BF first exceeded 3 (for at least 2 consecutive time points). C-D temporal
275 generalisation results. The left columns show classifier generalisation performance for the boat (C) and
276 star (D) between the different presentation durations. The right columns show corresponding
277 thresholded Bayes Factors (Yellow indicating above chance, and blue indicating below chance decoding).
278 Higher than chance generalisation (yellow) above the diagonal indicates slower processing of images in
279 the 5Hz condition relative to the 20Hz condition.
280

281 The temporal dynamics of target selection were revealed by decoding targets from non-targets. The
282 time-varying mean target-distractor decoding accuracy was computed separately for boat sequences
283 and star sequences (Figure 4). Target-distractor decoding performance peaked around 67% in the 5Hz
284 condition (Figure 4A), and around 60% in the 20Hz condition (Figure 4B). For both presentation rates,
285 peak decoding performance was around 500ms. In both conditions, decoding for star targets was above
286 chance earlier than for boats, which suggests that stars targets were easier to distinguish overall.
287 Decoding performance remained above chance for over 1000 ms in the 5Hz sequences, and for
288 approximately 800 ms in the 20Hz sequences.
289

290 The temporal generalisation approach revealed target selection was very similar between the 5Hz and
291 20Hz sequences. For both boat and star target sequences, the onset of target decoding occurred around
292 the same time, and cross-decoding was most evident along the diagonal, suggesting that target
293 selection processes occurred at the same latencies regardless of the sequence speed and image
294 duration (Figure 4c-d).

295

296 **Decoding categorical contrasts of 200 stimuli**

297 In the 5Hz condition, we observed above chance decoding for all categorical levels (Figure 5, blue lines),
298 starting at 100ms after stimulus onset for the categorical levels, and earlier (80ms) at the image level.
299 This difference may be caused by decodable low-level visual features at the image level, which are

300 controlled for by the exemplar-cross-validation approach at the categorical levels (Carlson et al., 2013;
301 Grootswagers, Wardle, et al., 2017). These decoding onsets correspond well to the existing decoding
302 literature, which has reported onsets for various categories between 80ms and 100ms (Carlson et al.,
303 2013; Cichy et al., 2014; Kaneshiro et al., 2015). For the animacy level, the results showed three distinct
304 peaks in decoding performance (150, 200ms and 400ms). In contrast, peak decoding happened around
305 200ms for category and object decoding and 130ms for image decoding. For all categorical levels,
306 above-chance decoding was sustained until around 500ms. Note that at 500ms, there were already two
307 new stimuli presented.

308

309 In the 20Hz condition (Figure 5, green lines), we again observed above-chance decoding for all levels.
310 Notably, the onset of decoding was around the same time point as in the 5Hz condition and subsequent
311 decoding followed the same trajectory but diverged later in the time series (indicated by the bottom
312 row of Bayes factors). The overall peak decoding performance was lower, and the peak decoding time
313 points appeared earlier in the time series. Decoding for all comparisons except object decoding
314 remained above chance until around 300ms, which included five subsequent stimulus presentations.
315 There was no difference between distractor processing on boat target and star target trials ($BF < 1/10$)
316 for all categorical contrasts.

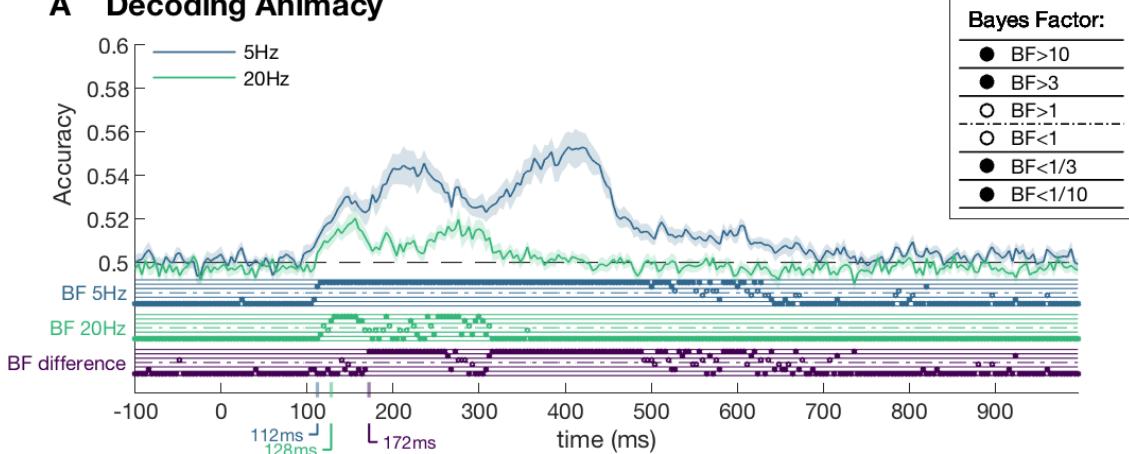
317

318 Temporal generalisation analyses were performed to compare categorical decoding between the 5Hz
319 and 20Hz conditions. For all three categorical levels, we observed similar onsets between presentation
320 durations, but longer subsequent processing for the 5Hz condition relative to the 20Hz condition (Figure
321 6). Notably, for the animacy distinction there was no evidence of generalisation between the 5Hz
322 sequence around 500-600ms and the 20Hz sequence at any time point, despite a difference between
323 decoding accuracies during this time period (as was seen in Figure 5). This suggests that a high-level
324 animacy-related process was present in the 5Hz condition but absent in the 20Hz condition. The

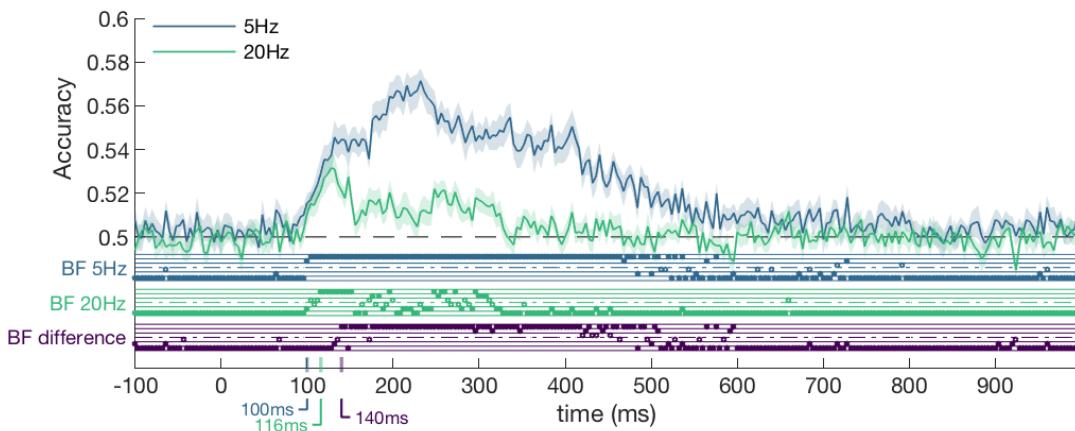
325 temporal generalisation analyses also showed consistent below chance generalisation between the
326 early and late responses. This phenomenon is consistent with previous decoding studies on visual object
327 categorisation (Carlson et al., 2013; Cichy et al., 2014), and has been suggested to be caused by the
328 stimulus offset, or by an adaptation or inhibition signal (Carlson et al., 2011, 2013; Contini et al., 2017).

329

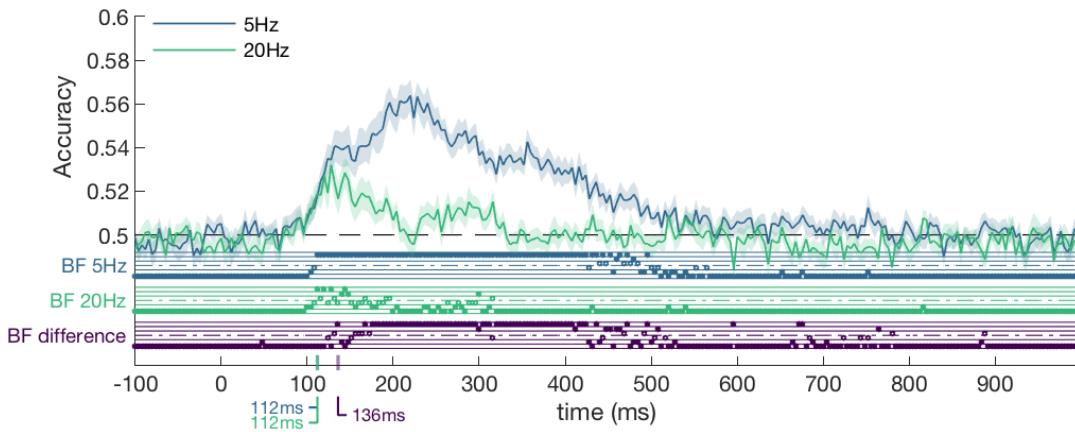
A Decoding Animacy



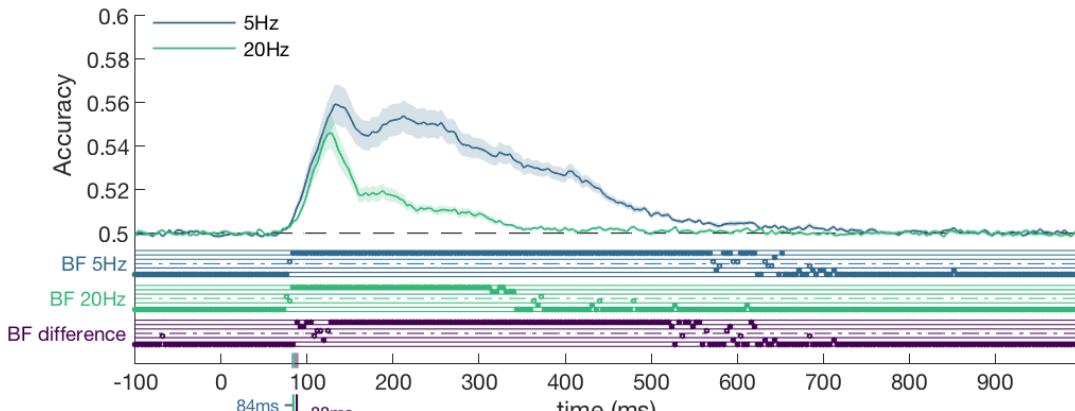
B Decoding Category



C Decoding Object

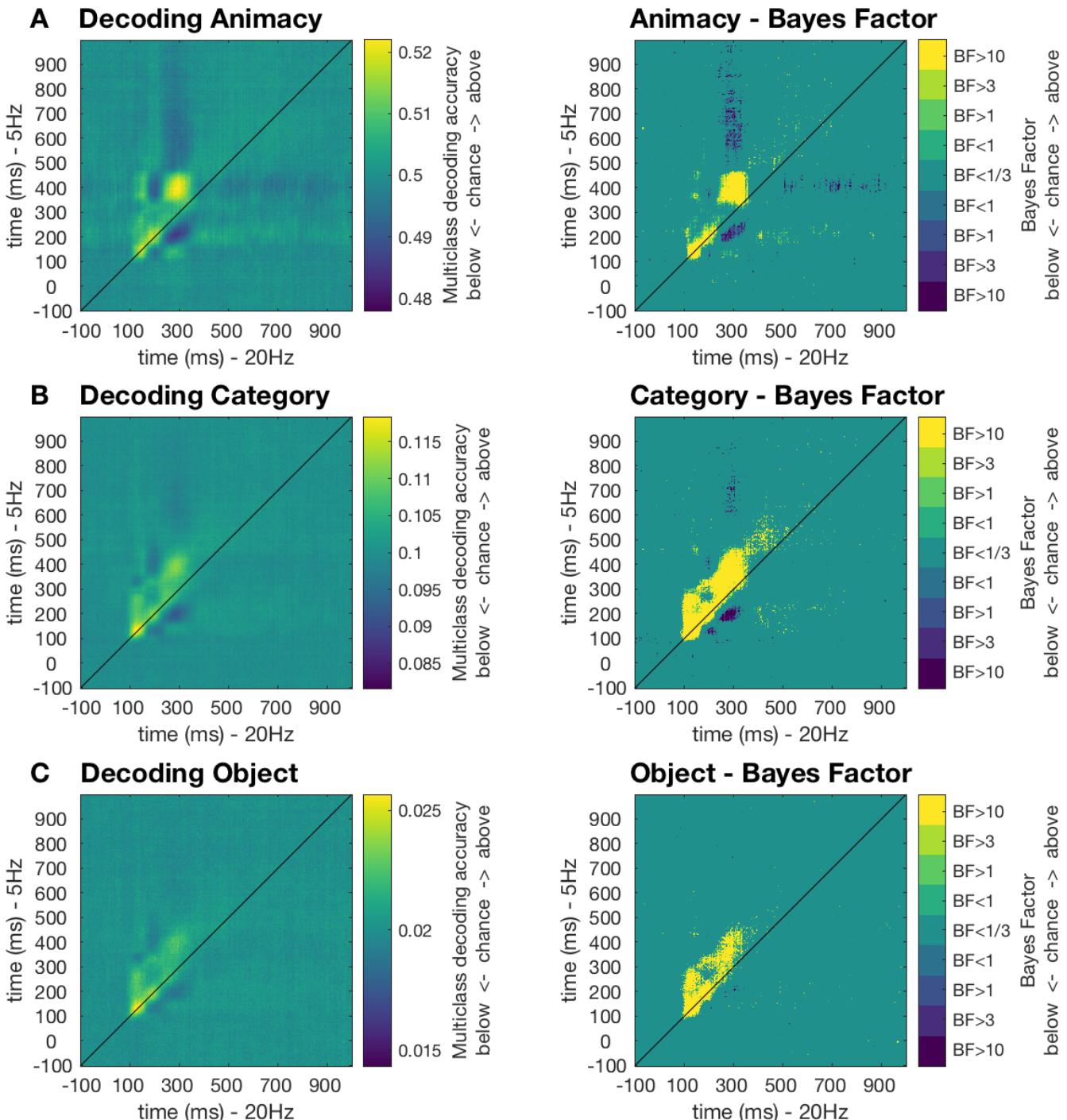


D Decoding Image



331 Figure 5. Mean decoding accuracy for 5Hz and 20Hz conditions. A) Decoding animacy (animate versus
332 inanimate). B) Mean pairwise decoding for the 10 categories (e.g., mammal, tools). C) Mean pairwise
333 decoding for 50 object categories (e.g., dog, giraffe). D) Mean pairwise decoding for all 200 images.
334 Shaded areas depict standard error of the mean across subjects. Dots below plots indicate thresholded
335 Bayes Factors (BF, see inset) for the 5Hz condition compared to chance (top rows), 20Hz condition
336 compared to chance (middle rows) and for the difference between the 5Hz and 20Hz results (bottom
337 rows). The time points where the BF first exceeded 3 are annotated below the x-axis.
338

339



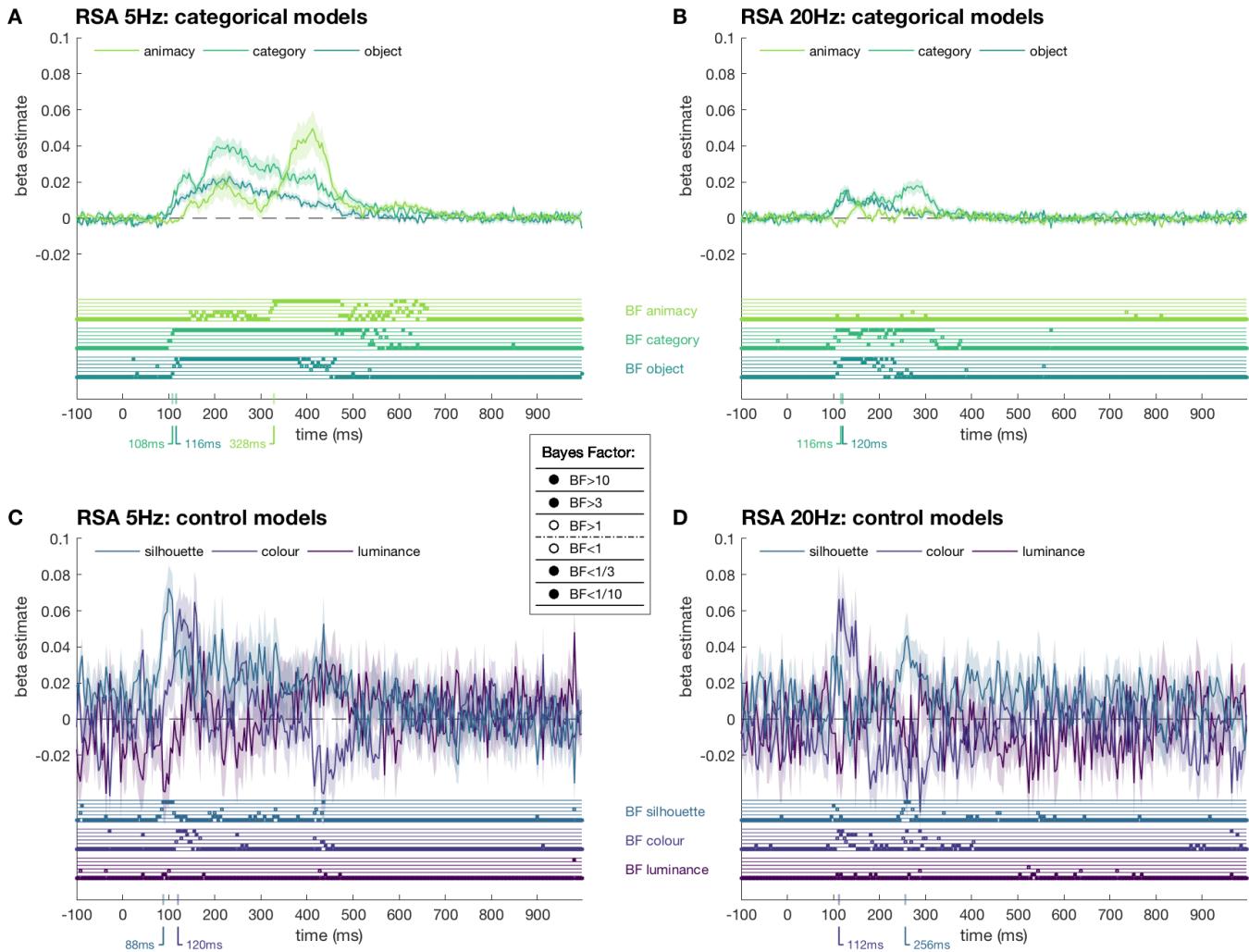
340

341 Figure 6. Temporal generalisation results. The left columns show classifier generalisation performance
342 for the three categorical levels (A-C) between the different presentation durations. The right columns
343 show corresponding thresholded Bayes Factors (Yellow indicating above chance, and blue indicating
344 below chance decoding). Higher than chance generalisation (yellow) above the diagonal indicates slower
345 processing in the 5Hz condition relative to the 20Hz condition.
346

347 **Representational dynamics of 200 stimuli**

348 Emerging representational structures of the 200 stimuli were studied in the Representational Similarity
349 Analysis (RSA) framework (Kriegeskorte & Kievit, 2013; Kriegeskorte, Mur, & Bandettini, 2008;
350 Kriegeskorte, Mur, Ruff, et al., 2008). A neural representational dissimilarity matrix (neural RDM) was
351 created for each subject, and each time point containing the dissimilarities between all 200 stimuli.
352 Neural RDMs were modelled as a linear combination of six candidate models; low-level image
353 silhouette, colour and luminance models, and one model for each of the three categorical levels. We
354 then analysed the mean beta estimates of the candidate models (Figure 7). For both presentation rates,
355 the silhouette model captured the early response in the data, followed by the colour, object, and
356 category models. These results quantify the contribution of low-level visual features to neural
357 dissimilarities. While low-level features were represented early in the signal, the categorical models also
358 explained unique variance in the data. In the 5Hz condition, the animacy model emerged last, while in
359 the 20Hz sequences the animacy model did not explain variance in the neural RDM at any time point. To
360 visualise and qualitatively explore the dynamic representational structure, we created a two-
361 dimensional embedding of all 200 images at three time points (200, 300, and 400ms), where the
362 distance between images reflects their mean dissimilarity across subjects (Figure 8; see supplementary
363 material for the corresponding neural RDMs at these time points).

364



365

366 Figure 7. RSA model tests. The neural RDMs of each subject were modelled as linear combination of six
367 candidate models: three categorical models (A&B), and three image feature models (C&D), separately
368 for the 5Hz sequences (A&C), and for the 20Hz sequences (B&D). Lines show estimated betas for the
369 models. Shaded areas reflect the standard error across subjects. Dots below plots indicate the
370 thresholded Bayes Factors (BF, see inset) for each beta estimate. Annotated below the x-axis are the
371 time points where the BF first exceeded 3 (for at least 2 consecutive time points).

372

A RSA 5Hz - MDS at 200ms



B RSA 20Hz - MDS at 200ms

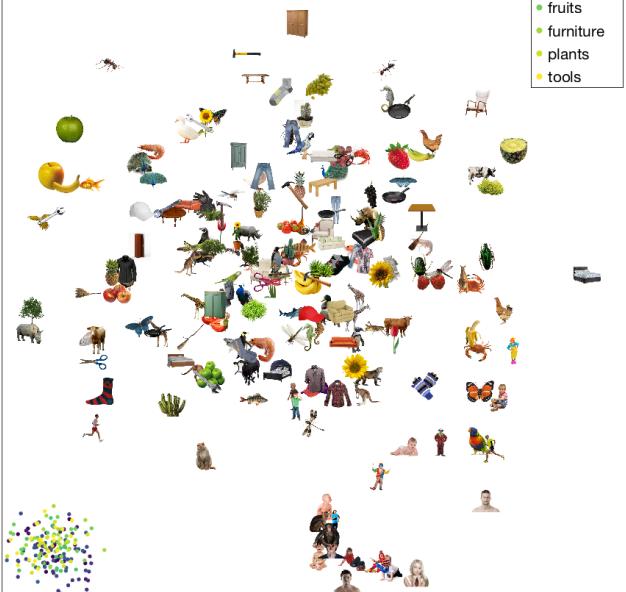


- aquatic
- bird
- human
- insect
- mammal
- clothing
- fruits
- furniture
- plants
- tools

C RSA 5Hz - MDS at 300ms



D RSA 20Hz - MDS at 300ms



E RSA 5Hz - MDS at 400ms



F RSA 20Hz - MDS at 400ms



374 Figure 8. Representational structure at three time points. Stimuli are embedded in a two-dimensional
375 space that reflects their pairwise distances. Stimuli that are shown further apart in this representation
376 evoked more dissimilar neural responses. The rows show three time points in the response, and the left
377 and right columns show data for the 5Hz and 20Hz conditions respectively. In the bottom left corner of
378 each plot, the same arrangement is shown, with images represented by dots coloured according to the
379 10 categories (see inset).

380

381 Discussion

382 In the current study, we characterised the representational dynamics of a large number of images in fast
383 presentation sequences. Previous work has used MEG and EEG decoding to investigate representations
384 of much smaller image sets using slow image presentation paradigms (Carlson et al., 2013; Cichy et al.,
385 2014; Contini et al., 2017; Grootswagers, Ritchie, et al., 2017; Kaiser et al., 2016; Kaneshiro et al., 2015;
386 Proklova et al., 2017; Ritchie, Tovar, & Carlson, 2015; Simanova, van Gerven, Oostenveld, & Hagoort,
387 2010); here we extend this work by looking at the representations of 200 objects during RSVP using
388 standard 64-channel EEG. For 5Hz and 20Hz sequences, all 200 images could be decoded at four
389 different categorical levels. Furthermore, neural responses to targets were distinct from those to
390 distractor stimuli. Above-chance decoding outlasted subsequent image presentations, supporting the
391 idea that multiple object representations can co-exist in the visual system at different stages of
392 processing (Marti & Dehaene, 2017). In keeping with the known hierarchical nature of the visual system,
393 RSA model testing suggested neural responses relied on low-level visual features early in the time series,
394 and subsequent processing was associated with increasing category abstraction (Carlson et al., 2013;
395 Cichy et al., 2014). Overall, we show the unprecedented ability of the human brain to process images
396 when pushing the limits of temporal perception.

397

398 Target decoding results revealed that neural responses to distractors diverged from star target
399 responses much earlier than boat targets. This supports our initial hypothesis that star targets would be
400 distinct from other images based on low-level visual features, unlike boat targets. The behavioural

401 results, however, revealed target detection did not differ across boat and star trials, indicating that
402 there was no “pop-out” effect of stars. This is despite anecdotal reports that participants found the star
403 targets easier. Target versus distractor decoding for boats and stars peaked at 500ms, supporting
404 previous evidence that high level cognitive processes mediate temporal selection (Marti & Dehaene,
405 2017; Sergent, Baillet, & Dehaene, 2005). These results suggest that distinguishable low-level features
406 do not help with target detection in RSVP sequences, at least in the current design with such high
407 variation in distractor images.

408

409 Target processing did not differ markedly across the different experimental durations. In both the 5Hz
410 and 20Hz sequences, targets could be distinguished from distractors for a long period of time, but this
411 was exaggerated for the 5Hz condition, where decoding was above chance for over 1000ms, compared
412 to 800ms in the 20Hz condition. Decoding was also higher in the 5Hz condition relative to the 20Hz
413 condition, but the dynamics of temporal selection processes were largely the same. The time of peak
414 decoding (500ms) was the same for both conditions, and time generalisation analyses revealed neural
415 processes occurred at the same latency in both conditions. This suggests that processes of target
416 selection are largely the same regardless of image presentation duration and frequency. Notably, target
417 processing was much more prolonged than categorical decoding for distractors, again indicative of
418 higher level cognitive processes at play for target detection. Note that the current experimental design
419 did not allow us to see which targets in the stream were missed, but effects are likely to be amplified for
420 correctly detected targets. Indeed, Marti & Dehaene (2017) found that late responses were sustained
421 for reported stimuli. Taken together, our results show that late target-related responses do not differ
422 dramatically in faster sequences relative to slower sequences.

423

424 Neural responses to the 200 non-task-relevant (distractor) objects are indicative of fairly automatic early
425 visual processing and divergence at later processing stages according to image duration. For all

426 contrasts, image presentation duration and cognitive task set did not influence the earliest processing
427 stages. When looking at decoding for the durations separately, onsets seemed to be earlier for the 5Hz
428 than 20Hz conditions, in accordance with recent work showing earlier onsets for longer image durations
429 (Mohsenzadeh et al., 2018). It is important to note, however, that higher signal strengths can also lead
430 to earlier decoding onsets (Grootswagers, Wardle, et al., 2017), thus differences between onsets must
431 be interpreted with caution in the context of larger peak decoding. Crucially, here Bayes factors
432 revealed evidence for no difference in decoding at these early time points between the 5Hz and 20Hz
433 image sequences (<150ms from image onset). Results from the temporal generalisation approach
434 supported this view, by showing that initial processing stages occurred at the same time for the 5Hz and
435 20Hz sequences, as seen by the above-chance decoding on the diagonal in Figure 5. Finally, for the three
436 categorical levels (animacy, category and object), Bayesian analyses revealed distractor processing did
437 not differ between boat and star trials. These results suggest that initial neural responses to all visual
438 stimuli were similar regardless of their presentation duration.

439

440 Previous work has shown that, using MEG, it is possible to use decoding to investigate target-related
441 processes in RSVP streams (Marti & Dehaene, 2017; Mohsenzadeh et al., 2018a). For example,
442 Mohsenzadeh et al. used 306-channel MEG to decode 12 target faces from 12 non-target objects in
443 RSVP streams, analysing only the middle image in the stream to study feedforward versus feedback
444 processes. As part of a study investigating temporal selection mechanisms, Marti & Dehaene showed
445 that a classifier trained on 5 categories using a separate localiser could generalise to distractor items
446 around the target. In contrast to these studies, here we decoded object representations using a 64-
447 channel EEG, a much larger set of images (200) in a sequence, and no separate localiser. The results
448 from our approach also corroborated previous work decoding the representations of objects presented
449 in isolation (Carlson et al., 2013; Cichy et al., 2014; Kaneshiro et al., 2015). Our results showed that
450 decoding objects in RSVP streams have similar decoding onsets as previously reported (Carlson et al.,

451 2013; Cichy et al., 2014; Kaneshiro et al., 2015). This validates the RSVP approach as a method to study
452 representational dynamics. We further found that the 20Hz condition limited visual processing
453 compared to 5Hz, which shows that this paradigm can be utilised to bias the extent of visual processing
454 at different image presentation rates. In sum, our results confirm that long ISIs are not necessary for
455 multivariate analyses. This thus allows analysing all presentations in an RSVP sequence, rather than
456 limiting the scope to selected presentations (e.g., targets) in the streams. Here we have demonstrated
457 the potential by studying the representational dynamics of 200 objects in one short EEG session. Future
458 work can adopt similar approaches to investigate for example prediction, priming, masking, or
459 attentional effects on the processing of distractors in RSVP sequences.

460

461 Despite similar early processing stages, later processing diverged according to image presentation
462 duration. Representations during 5Hz sequences were stronger and lasted longer than those during
463 20Hz sequences, and temporal generalisation analyses showed that processes were prolonged for the
464 5Hz relative to the 20Hz condition. It could be that longer image durations allow more consolidation,
465 potentially due to recurrent processing. It is also possible that longer durations allow time to reach
466 some kind of threshold, which triggers further processing. Note that image duration and ISI are
467 conflated in this design, so we cannot conclude whether or if stronger and longer processing occurs due
468 to longer image presentation or due to delayed masking from the next stimulus. Future work can build
469 on this approach to investigate the temporal limits of visual perception.

470

471 The RSA regression analyses provided insight into the differences in processing between the 5Hz and
472 20Hz sequences. The category decoding analyses were performed using a leave-one-exemplar out cross-
473 validation approach, which means that the classifier always had to generalise to new images, reducing
474 the likelihood that low-level features would drive the results. However, there can still be consistent low-
475 level features between the categories that can contribute to classification. The regression RSA technique

476 aimed to dissociate the unique contributions of each of the categorical and low-level featural models. In
477 accordance with the decoding results, processes early in the time series (~100-150ms) were mostly
478 explained by the low-level silhouette model and then the colour model for the 5Hz and 20Hz conditions
479 (Carlson et al., 2011). Subsequent processing, however, elucidated the differential contributions of the
480 different categorical contrasts, and how this varied for the different image durations. For the 5Hz
481 condition, the category model appeared to have the largest unique contribution around 200ms, and the
482 animacy model accounted for the most variance at about 400ms, indicating that increasing category
483 abstraction occurred at higher levels of visual processing (Carlson et al., 2013; Cichy et al., 2014; Contini
484 et al., 2017; Kriegeskorte, Mur, Ruff, et al., 2008). In contrast, the animacy model had no unique
485 contribution to the signal for the in 20Hz sequences. The time course of the animacy model regression
486 for the 5Hz condition (>350ms) suggests that the animate-inanimate difference might exclusively
487 account for the prolonged decoding in the 5Hz condition relative to the 20Hz condition. This could imply
488 that a high-level animacy effect requires sufficient evidence accumulation to proceed, which does not
489 happen at 20Hz presentation rates. The finding that longer image presentations allow higher level
490 processing is supported by steady-state visual evoked potential (SSVEP) work showing that images
491 presented at faster frequencies are biased towards earlier visual processes in contrast to slower
492 frequencies which allow higher level processing (Collins, Robinson, & Behrmann, 2018).

493

494 When qualitatively inspecting the visualisation of the representational structure (Figure 8), we noticed a
495 clear animate versus inanimate organisation in the 5Hz presentation condition. At 200ms in the
496 response, the structure reflected mostly natural versus artificial, with plants, fruits and animals all
497 clustering on one side. In line with the decoding and RSA results, the structure at 400ms showed a clear
498 animate – inanimate distinction (Caramazza & Shelton, 1998), which is commonly observed in neural
499 responses in the ventral temporal cortex (Cichy et al., 2014; Konkle & Caramazza, 2013; Kriegeskorte,
500 Mur, Ruff, et al., 2008; Proklova, Kaiser, & Peelen, 2016) and has been shown to match human

501 categorisation behaviour well (Bracci & Op de Beeck, 2016; Carlson, Ritchie, Kriegeskorte, Durvasula, &
502 Ma, 2014; Grootswagers, Cichy, & Carlson, 2018; Mur et al., 2013; Ritchie, Tovar, & Carlson, 2015). In
503 the animate – inanimate organisation primates were located at the far end of the animate side, which
504 may reflect a continuum of biological classes in the brain (Connolly et al., 2012; Sha et al., 2015) or
505 typicality (Grootswagers, Ritchie, et al., 2017; Iordan, Greene, Beck, & Fei-Fei, 2016; Posner & Keele,
506 1968; E. H. Rosch, 1973; E. Rosch & Mervis, 1975). No animacy structure was apparent for the 20Hz
507 condition (as evidenced by the RSA results), but rather individual categorical clusters seem to have
508 emerged (in line with the RSA results), such as human faces, and, later, humans and primates as a
509 category. Interestingly, in these visualisations gloves were grouped with humans and primates, which
510 could mean they were perceived as body parts, rather than inanimate objects. While these
511 visualisations allow for such qualitative speculation, the quantitative RSA modelling results highlight the
512 level of detail in the representation structure that can be obtained using EEG decoding and fast
513 presentation rates. Here, we used a common 64-channel EEG, but future work can use this approach in
514 combination with high-density EEG or other neuroimaging methods that are sensitive to finer spatial
515 patterns, such as MEG.

516

517 One remaining question is the role that low-level image statistics play in our results. The RSA approach
518 showed that low-level control models explained early neural responses to the stimuli. The current
519 stimulus set consisted of segmented coloured objects, which were not matched on low-level features
520 such as colour, orientation, shape, and size. Future work can build on the current paradigm using a
521 stimulus set that for example contains orthogonal shape and category dimensions (Bracci, Kalfas, & Op
522 de Beeck, 2017; Bracci & Op de Beeck, 2016; Proklova et al., 2017, 2016), or test the decodability of
523 these features using for example texture stimuli with similar features (Long, Konkle, Cohen, & Alvarez,
524 2016; Long, Yu, & Konkle, 2017). Such extensions can help unravel the relationship between object

525 features and categories, and increase our understanding of how this inherent relationship guides
526 categorical abstraction in the visual system.

527

528 In conclusion, our results show that we can study the representational dynamics of more than 200
529 objects in one short EEG session. We were able to characterise the time courses of multiple categorical
530 contrasts from the same images, indicating that all objects reached abstract categorical stages of
531 perception despite being presented for short durations. Here, we took advantage of the high temporal
532 resolution of both the human visual system and common neuroimaging techniques such as EEG and
533 MEG. These results confirm that long ISIs are not necessary for multivariate analyses, as they do not
534 require a resting baseline as in ERP analyses. Thus, future MVPA studies on visual perception should
535 consider using fast presentation rates as this allows for a substantial increase of the number of
536 presentations, stimuli, or experimental conditions. This offers unprecedented potential for studying the
537 temporal dynamics of visual perception and attention.

538

539 **Acknowledgements**

540 We thank Nick McNair for assisting with data collection. This research was supported by an Australian
541 Research Council Future Fellowship (FT120100816) and an Australian Research Council Discovery
542 project (DP160101300) awarded to T.A.C. The authors acknowledge the University of Sydney HPC
543 service for providing High Performance Computing resources. The authors declare no competing
544 financial interests.

545

546 References

547

- 548 Bracci, S., Kalfas, I., & Op de Beeck, H. P. (2017). The ventral visual pathway represents animal
549 appearance over animacy, unlike human behavior and deep neural networks. *BioRxiv*, 228932.
550 <https://doi.org/10.1101/228932>

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

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

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

560 Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution
561 decoding of object position and category. *Journal of Vision*, 11(10), 9.
562 <https://doi.org/10.1167/11.10.9>

563 Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction Time for Object
564 Categorization Is Predicted by Representational Distance. *Journal of Cognitive Neuroscience*,
565 26(1), 132–142. https://doi.org/10.1162/jocn_a_00476

566 Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object
567 vision: The first 1000 ms. *Journal of Vision*, 13(10), 1. <https://doi.org/10.1167/13.10.1>

568 Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time.
569 *Nature Neuroscience*, 17(3), 455–462. <https://doi.org/10.1038/nn.3635>

- 570 Collins, E., Robinson, A. K., & Behrmann, M. (2018). Distinct neural processes for the perception of
571 familiar versus unfamiliar faces along the visual hierarchy revealed by EEG. *NeuroImage*, 181,
572 120–131. <https://doi.org/10.1016/j.neuroimage.2018.06.080>
- 573 Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., ... Haxby, J. V. (2012).
574 The Representation of Biological Classes in the Human Brain. *The Journal of Neuroscience*, 32(8),
575 2608–2618. <https://doi.org/10.1523/JNEUROSCI.5547-11.2012>
- 576 Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in
577 the human brain: From visual features to categorical decisions. *Neuropsychologia*, 105, 165–176.
578 <https://doi.org/10.1016/j.neuropsychologia.2017.02.013>
- 579 Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100
580 ms. *Journal of Vision*, 10(4), 16. <https://doi.org/10.1167/10.4.16>
- 581 Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG
582 dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1),
583 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- 584 Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on
585 Psychological Science*, 6(3), 274–290. <https://doi.org/10.1177/1745691611406920>
- 586 Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and
587 its role in categorization. *Nature Reviews Neuroscience*, 15(8), 536–548.
588 <https://doi.org/10.1038/nrn3747>
- 589 Grootswagers, T., Cichy, R. M., & Carlson, T. A. (2018). Finding decodable information that can be read
590 out in behaviour. *NeuroImage*, 179, 252–262.
591 <https://doi.org/10.1016/j.neuroimage.2018.06.022>
- 592 Grootswagers, T., Ritchie, J. B., Wardle, S. G., Heathcote, A., & Carlson, T. A. (2017). Asymmetric
593 Compression of Representational Space for Object Animacy Categorization under Degraded

- 594 Viewing Conditions. *Journal of Cognitive Neuroscience*, 29(12), 1995–2010.
- 595 https://doi.org/10.1162/jocn_a_01177
- 596 Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked
597 Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging
598 Data. *Journal of Cognitive Neuroscience*, 29(4), 677–697. https://doi.org/10.1162/jocn_a_01068
- 599 Haynes, J.-D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and
600 Perspectives. *Neuron*, 87(2), 257–270. <https://doi.org/10.1016/j.neuron.2015.05.025>
- 601 Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2016). Typicality sharpens category
602 representations in object-selective cortex. *NeuroImage*, 134, 170–179.
603 <https://doi.org/10.1016/j.neuroimage.2016.04.012>
- 604 Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in
605 the human visual system. *Journal of Neurophysiology*, 111(1), 91–102.
606 <https://doi.org/10.1152/jn.00394.2013>
- 607 Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.
- 608 Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed
609 by MEG and fMRI decoding. *Journal of Neurophysiology*, 115(4), 2246–2250.
610 <https://doi.org/10.1152/jn.01074.2015>
- 611 Kaneshiro, B., Guimaraes, M. P., Kim, H.-S., Norcia, A. M., & Suppes, P. (2015). A Representational
612 Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification.
613 *PLOS ONE*, 10(8), e0135697. <https://doi.org/10.1371/journal.pone.0135697>
- 614 Keysers, C., Xiao, D.-K., Földiák, P., & Perrett, D. I. (2001). The Speed of Sight. *Journal of Cognitive
615 Neuroscience*, 13(1), 90–101. <https://doi.org/10.1162/089892901564199>
- 616 Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object Category Structure in Response Patterns of
617 Neuronal Population in Monkey Inferior Temporal Cortex. *Journal of Neurophysiology*, 97(6),
618 4296–4309. <https://doi.org/10.1152/jn.00024.2007>

- 619 King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal
620 generalization method. *Trends in Cognitive Sciences*, 18(4), 203–210.
621 <https://doi.org/10.1016/j.tics.2014.01.002>
- 622 Konkle, T., & Caramazza, A. (2013). Tripartite Organization of the Ventral Stream by Animacy and Object
623 Size. *Journal of Neuroscience*, 33(25), 10235–10242. <https://doi.org/10.1523/JNEUROSCI.0983-13.2013>
- 625 Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation,
626 and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412.
627 <https://doi.org/10.1016/j.tics.2013.06.007>
- 628 Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational Similarity Analysis - Connecting
629 the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience*, 2, 4.
630 <https://doi.org/10.3389/neuro.06.004.2008>
- 631 Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008).
632 Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey.
633 *Neuron*, 60(6), 1126–1141. <https://doi.org/10.1016/j.neuron.2008.10.043>
- 634 Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features distinguish
635 objects of different real-world sizes. *Journal of Experimental Psychology: General*, 145(1), 95.
- 636 Long, B., Yu, C.-P., & Konkle, T. (2017). A mid-level organization of the ventral stream. *BioRxiv*, 213934.
637 <https://doi.org/10.1101/213934>
- 638 Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning
639 Research*, 9(Nov), 2579–2605.
- 640 Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural
641 scenes: how robust at very low contrasts? *European Journal of Neuroscience*, 21(7), 2007–2018.
642 <https://doi.org/10.1111/j.1460-9568.2005.04029.x>

- 643 Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization:
644 Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*,
645 15(1), 28–35. <https://doi.org/10.3758/PBR.15.1.28>
- 646 Mack, M. L., & Palmeri, T. J. (2011). The Timing of Visual Object Categorization. *Frontiers in Psychology*,
647 2. <https://doi.org/10.3389/fpsyg.2011.00165>
- 648 Mack, M. L., & Palmeri, T. J. (2015). The Dynamics of Categorization: Unraveling Rapid Categorization.
649 *Journal of Experimental Psychology: General*, No Pagination Specified.
650 <https://doi.org/10.1037/a0039184>
- 651 Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain?
652 *Trends in Cognitive Sciences*, 15(3), 97–103. <https://doi.org/10.1016/j.tics.2011.01.004>
- 653 Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid
654 visual streams. *Nature Communications*, 8(1), 1955. <https://doi.org/10.1038/s41467-017-02079-x>
- 655 x
- 656 Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic Population
657 Coding of Category Information in Inferior Temporal and Prefrontal Cortex. *Journal of
658 Neurophysiology*, 100(3), 1407–1419. <https://doi.org/10.1152/jn.90248.2008>
- 659 Mohsenzadeh, Y., Qin, S., Cichy, R. M., & Pantazis, D. (2018). Ultra-Rapid serial visual presentation
660 reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *eLife*, 7,
661 e36329. <https://doi.org/10.7554/eLife.36329>
- 662 Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human Object-
663 Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. *Frontiers in
664 Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00128>
- 665 Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and
666 ERP measurements. *Clinical Neurophysiology*, 112(4), 713–719. [https://doi.org/10.1016/S1388-2457\(00\)00527-7](https://doi.org/10.1016/S1388-2457(00)00527-7)

- 668 Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern
669 Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, 10.
670 <https://doi.org/10.3389/fninf.2016.00027>
- 671 Peelen, M. V., & Caramazza, A. (2012). Conceptual Object Representations in Human Anterior Temporal
672 Cortex. *Journal of Neuroscience*, 32(45), 15728–15736.
673 <https://doi.org/10.1523/JNEUROSCI.1953-12.2012>
- 674 Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental
675 Psychology*, 77(3p1), 353. <http://dx.doi.org/10.1037/h0025953>
- 676 Potter, M. C. (1975). Meaning in visual search. *Science*, 187(4180), 965–966.
- 677 Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology:
678 Human Learning and Memory*, 2(5), 509.
- 679 Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms
680 per picture. *Attention, Perception, & Psychophysics*, 76(2), 270–279.
- 681 Proklova, D., Kaiser, D., & Peelen, M. (2017). MEG sensor patterns reflect perceptual but not categorical
682 similarity of animate and inanimate objects. *BioRxiv*, 238584. <https://doi.org/10.1101/238584>
- 683 Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and
684 Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. *Journal of
685 Cognitive Neuroscience*, 1–13. https://doi.org/10.1162/jocn_a_00924
- 686 Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging Object Representations in the Visual System
687 Predict Reaction Times for Categorization. *PLoS Comput Biol*, 11(6), e1004316.
688 <https://doi.org/10.1371/journal.pcbi.1004316>
- 689 Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, 4(3), 328–350.
690 [https://doi.org/10.1016/0010-0285\(73\)90017-0](https://doi.org/10.1016/0010-0285(73)90017-0)
- 691 Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories.
692 *Cognitive Psychology*, 7(4), 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9)

- 693 Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images
694 reveals a robust face-selective electrophysiological response in the human brain. *Journal of*
695 *Vision*, 15(1), 18–18. <https://doi.org/10.1167/15.1.18>
- 696 Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting
697 and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237.
- 698 Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization
699 of natural images. *Nature Neuroscience*, 5(7), 629–630. <https://doi.org/10.1038/nn866>
- 700 Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to
701 consciousness during the attentional blink. *Nature Neuroscience*, 8(10), 1391.
- 702 Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C.
703 (2015). The Animacy Continuum in the Human Ventral Vision Pathway. *Journal of Cognitive*
704 *Neuroscience*, 27(4), 665–678. https://doi.org/10.1162/jocn_a_00733
- 705 Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying Object Categories from
706 Event-Related EEG: Toward Decoding of Conceptual Representations. *PLoS ONE*, 5(12), e14465.
707 <https://doi.org/10.1371/journal.pone.0014465>
- 708 Teichmann, L., Grootswagers, T., Carlson, T., & Rich, A. N. (2018). Decoding Digits and Dice with
709 Magnetoencephalography: Evidence for a Shared Representation of Magnitude. *Journal of*
710 *Cognitive Neuroscience*, 30(7), 999–1010. https://doi.org/10.1162/jocn_a_01257
- 711 Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*,
712 381(6582), 520–522. <https://doi.org/doi:10.1038/381520a0>
- 713 VanRullen, R., & Thorpe, S. J. (2001). The Time Course of Visual Processing: From Early Perception to
714 Decision-Making. *Journal of Cognitive Neuroscience*, 13(4), 454–461.
715 <https://doi.org/10.1162/08989290152001880>
- 716 Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic*
717 *Bulletin & Review*, 14(5), 779–804. <https://doi.org/10.3758/BF03194105>

- 718 Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S.-M., & Carlson, T. A. (2016).
- 719 Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured
- 720 with MEG. *NeuroImage*, 132, 59–70. <https://doi.org/10.1016/j.neuroimage.2016.02.019>
- 721 Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical
- 722 Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives*
- 723 on Psychological Science, 6(3), 291–298. <https://doi.org/10.1177/1745691611406923>
- 724