bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Tijl Grootswagers*[a,b,c,1], Amanda K. Robinson*[a,b,c,2], Thomas A. Carlson[a,b,3]

*equal contribution

[a] School of Psychology, University of Sydney, NSW, 2006, Australia

[b] ARC Centre of Excellence in Cognition and Its Disorders, NSW, 2109, Australia

[c] Department of Cognitive Science, Macquarie University, NSW, 2109, Australia

[1] tijl.grootswagers@sydney.edu.au

[2] amanda.robinson@sydney.edu.au

[3] thomas.carlson@sydney.edu.au

Abstract

In our daily lives, we are bombarded with a stream of rapidly changing visual input. Humans have the remarkable capacity to detect and identify objects in fast-changing scenes. Yet, when studying brain representations, stimuli are generally presented in isolation. Here, we studied the dynamics of human vision using a combination of fast stimulus presentation rates, electroencephalography and multivariate decoding analyses. Using a rapid presentation rate of 5 images per second, we obtained the representational structure of a large number of stimuli, and showed the emerging abstract categorical organisation of this structure. In a second experiment, we replicated these results using an ultra-rapid presentation rate of 20 images per second. Our results show that the combination of naturalistic stimulus presentation rates and multivariate decoding analyses has unprecedented potential for studying the temporal dynamics of the structure of representations in the human visual system.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

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

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

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

their neural representations can co-exist in the visual system (Marti & Dehaene, 2017; Mohsenzadeh, Qin, Cichy, & Pantazis, 2018; Rossion, Torfs, Jacques, & Liu-Shuang, 2015; Rousselet, Fabre-Thorpe, & Thorpe, 2002). Behavioural work has additionally shown that the human visual system can extract abstract information from a visual stimulus at very fast presentation rates (Crouzet, Kirchner, & Thorpe, 2010; Keysers et al., 2001; Macé, Thorpe, & Fabre-Thorpe, 2005; Mack et al., 2008; Mack & Palmeri, 2015; Marti & Dehaene, 2017; Potter, 1975, 1976; Potter et al., 2014; Rossion et al., 2015; Thorpe, Fize, & Marlot, 1996). In the current study, we draw on this human capacity and study visual object recognition using fast stimulus presentation rates and multivariate decoding analyses of EEG evoked responses (Grootswagers, Wardle, et al., 2017). We used a rapid serial visual presentation (RSVP) paradigm to study the representations of a large set of visual objects presented at a speed of 5 images per second (200ms per image; "rapid" condition). The objects were carefully selected to allow categorisation at three different levels. The high presentation rate enabled us to obtain 40 repetitions of 200 different stimuli in a short EEG session. We additionally examined the effect of higher level cognitive processes on the emerging representations by having participants detect targets that were identifiable based on low-level visual features or abstract categories in separate trials. We successfully decoded different categorical contrasts for the 200 objects, suggesting that individual stimuli were processed up to abstract categorical representations. Strikingly, we found similar results in a follow-up experimental session, where we used a much higher presentation rate of 20 images per second (50ms per image: "ultra-rapid" condition). The unprecedented ability to test such large numbers of different stimuli in relatively short scanning sessions shows great potential for studying the dynamics of the structure of information in the human visual system.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

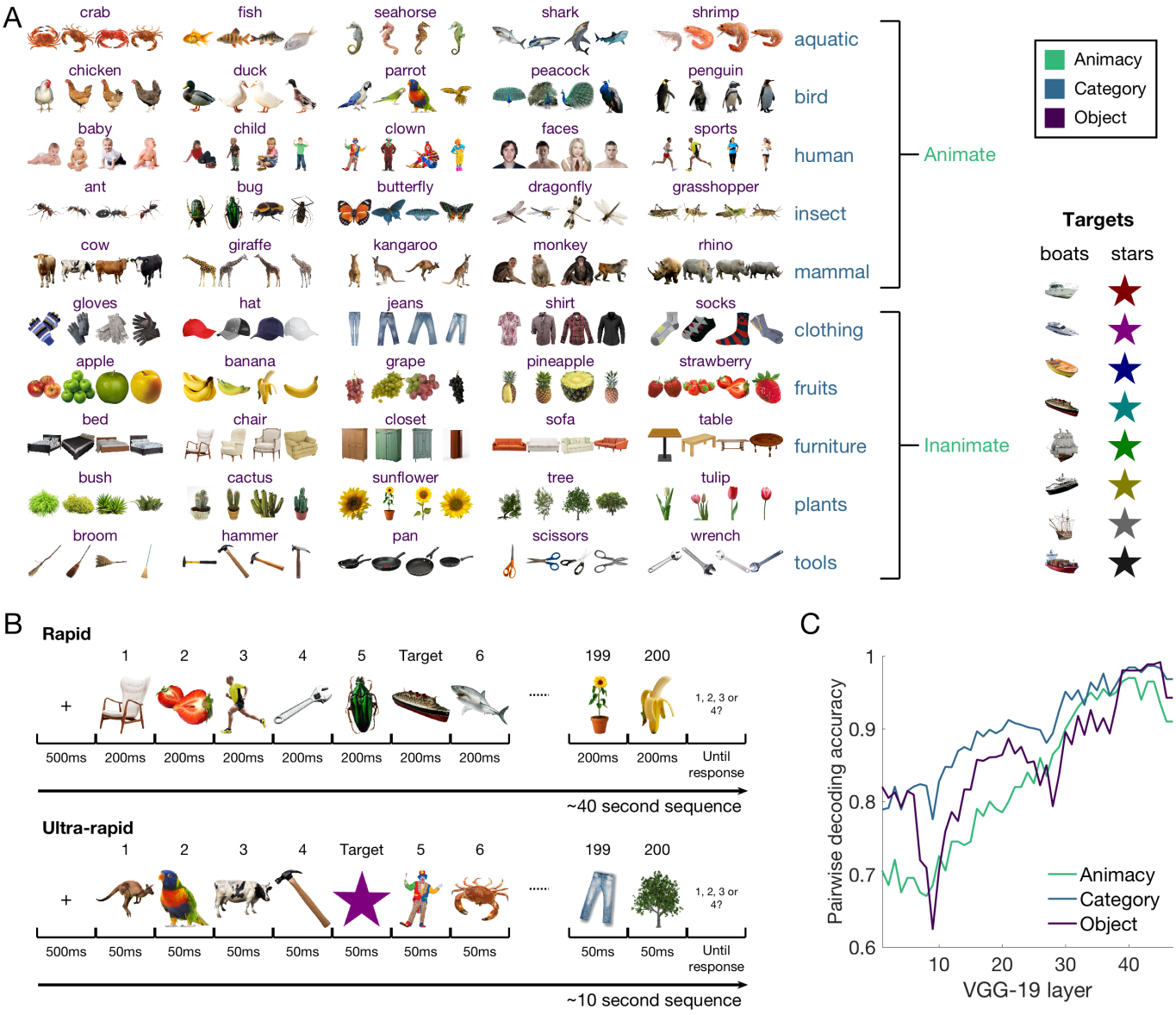


Figure 1. Stimuli and design. A) Experimental stimuli. There were 200 images of objects (obtained from www.pngimg.com), organised in categories at three different levels: Animacy (animate, inanimate), category (10 categories e.g., mammal, tool, flower) and object (50 categories e.g., cow, dog, giraffe). In the experiment, participants were asked to count the number of target objects from two categories: boats and geometric star shapes, each with eight images. B) Experimental design. Trials consisted of all 200 images presented in random order, with 1-4 targets interspersed throughout. Images were presented for 200ms each in session 1, and 50ms in session 2. C) Decoding the categorical levels from deep neural network activation profiles. For each layer in the VGG19-network (x-axis), we decoded the three categorical levels using a leave-one-exemplar-out cross-validation procedure. Plotted is the mean decoding accuracy (y-axis) for each layer of the network.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Stimuli

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

To predict the contribution of low level visual features to our results, we trained classifiers on activation profiles obtained from a pre-trained deep neural network (DNN), VGG-19 (Simonyan & Zisserman, 2014). This network was originally trained to classify objects into 1000 categories using 45 hidden layers. For each of our 200 experimental stimuli, we obtained the activation values of all units in each layer of the DNN. As there are many more units in a layer than we have stimuli, we applied principal component analysis (PCA) on the activation patters to retain the first 200 components without loss of information.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Then, for each layer, we applied a linear discriminant analysis classifier to the activation of units in this layer. The leave-one-exemplar-out cross-validated classification accuracy was calculated for the three categorical contrasts used in this study (Figure 1c), which indicated category decoding accuracy increased with deep neural network layer.

Participants and experimental procedure

Participants were 16 adults recruited from the University of Sydney (5 females; age range 18-38 years) in return for payment or course credit. The study was approved by the University of Sydney ethics committee and informed consent was obtained from all participants. Participants viewed 40 sequences of objects, each lasting between 40.2 - 40.8 seconds (depending on the number of targets in the sequence). In each sequence, the 200 stimuli were presented in random order, for a duration of 200ms each with no gap between successive images (5 images per second; “rapid” condition). In addition to the 200 stimuli, target stimuli were inserted throughout the sequence (Figure 1c). In half of the sequences, the target stimuli were boats, and in the other sequences, the target stimuli were geometric stars (Figure 1). A random number between 1 and 4 targets were presented in the sequence, with the condition that targets could not appear within the first 10 or last 10 images, and ensuring there were at least 12 non-target stimuli between subsequent targets. At the start of each sequence, participants were prompted to count the number of targets in the sequence (“Count the boats in the trial” or “Count the stars in the trial” in random order) and the 8 potential targets were shown. They were instructed to respond at the end of the sequence using a 4-way button box. After each sequence, participants received feedback. They started the next sequence with a button press. This session lasted approximately 40 minutes in total. After a short break, the second experimental session started, and participants performed another 40 sequences using the same procedure as session one, except that the images were presented for only 50ms (a presentation speed of 20 images per second; “ultra-rapid” condition). The second session lasted about 10 minutes.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

137

138 **EEG recordings and preprocessing**

139 Continuous EEG data were recorded using a BrainVision ActiChamp system, digitized at a 1000-Hz
140 sample rate. The 64 electrodes were arranged according to the international standard 10–10 system for
141 electrode placement (Oostenveld & Praamstra, 2001). During recording, all scalp electrodes were
142 referenced to Cz. Data were preprocessed offline using EEGlab (Delorme & Makeig, 2004). Data were
143 bandpass filtered between [0.1 – 100Hz] and downsampled to 250Hz. Epochs were then created for
144 each stimulus presentation (except targets) ranging from [-100 to 1000ms] relative to stimulus onset.
145 No further preprocessing steps were applied. For comparing targets versus distractors, we created
146 larger epochs ranging from [-500 to 3000ms] relative to the onset of a target, and for each target *t*, we
147 selected at random (to avoid temporal dependencies between target and distractor data) the *t*-4th, *t*-5th,
148 or *t*-6th preceding distractor and created a matching epoch relative to the onset of that distractor.
149 Choosing distractors in this way meant that the number of targets and distractors were matched per
150 sequence and that the neural representations of targets and distractors were unlikely to overlap.

151

152 **Decoding analysis**

153 A standard MVPA decoding pipeline (Grootswagers, Wardle, et al., 2017; Oosterhof, Connolly, & Haxby,
154 2016) was applied to each stimulus presentation epoch in the sequence to investigate object
155 representations in fast sequences. To investigate the temporal dynamics of target selection, we
156 compared neural responses to targets with those to non-target distractor stimuli. Classifiers were then
157 trained to distinguish targets from non-targets separately for the rapid and ultra-rapid sequences, and
158 for boat and star target sequences.

159

160 We investigated object representations for the 200 non-target images using multiple categorical
161 distinctions. First, we decoded three contrasts that impose different amounts of categorical abstraction.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

At the highest level, we decoded animacy (i.e., animate versus inanimate objects). The next contrast was the category tier (10 classes, e.g., mammal, insect, furniture, tool, etc.) where we decoded all 45 possible pairwise combinations. The lowest categorical level was the object level (50 classes, e.g., cow, butterfly, table, hammer, etc.). Here, we decoded all 1225 possible pairwise object combinations (i.e., cow versus butterfly, cow versus table, etc.). Finally, at the lowest level, we investigated image-level representations by decoding all 19900 possible pairwise combinations of the 200 stimuli.

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

For all contrasts, we used a leave-one-sequence-out cross-validation scheme in combination with a linear discriminant analysis classifier, and report the mean cross-validated decoding accuracy. All steps in the decoding analysis were implemented in CoSMoMVPA (Oosterhof et al., 2016).

Representational Similarity Analysis

To study the emerging representational structure of our 200 stimuli, we analysed our data using the Representational Similarity Analysis (RSA) framework (Kriegeskorte & Kievit, 2013; Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008), which allows comparing models of object representations. The decoding results at the image level were organised into a 200 by 200 neural representational dissimilarity matrix (RDM), which for each pair of images, contains the mean cross-

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

validated decoding accuracy (images that evoke more dissimilar neural responses are better decodable). One neural RDM was created for each subject, and each time point. The time-varying neural RDMs of each subject were then modelled as a linear combination of the candidate models using a GLM (Oosterhof et al., 2016; Proklova, Kaiser, & Peelen, 2017); For each time point, the lower triangles of the neural RDM and candidate models were vectorised, and regression coefficients were obtained for all candidate models. The candidate models consisted of a low-level image outline similarity model (which is a good predictor of differences in brain responses (Carlson et al., 2011; Teichmann, Grootswagers, Carlson, & Rich, 2018; Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016)), and one model for each of the three categorical levels, grouping images from the same category. This resulted in one beta estimate for each model, subject, and time point. We then analysed at the group level the mean beta estimates across subjects. To visualise the dynamic representational structure, at each point in time, we created a two-dimensional embedding of all 200 images, where the distance between images reflects their mean dissimilarity across subjects. To compute the two-dimensional embedding, we applied t-SNE (Maaten & Hinton, 2008) to the mean neural RDMs.

We investigated the stage-wise computations by comparing the neural RDMs with RDMs created from deep neural network layer activation patterns (Bankson, Hebart, Groen, & Baker, 2018; Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016; Greene & Hansen, 2018; Güçlü & Gerven, 2015; Khaligh-Razavi & Kriegeskorte, 2014). For each of the 200 stimuli we used the PCA-transformed activation profiles obtained from all units a pre-trained deep neural network (DNN), VGG-19 (Simonyan & Zisserman, 2014). Then, for each pair of stimuli, we computed the Euclidean distance between activation patterns, resulting in one RDM for each layer of the deep neural network. The deep neural network RDMs for each layer were correlated to each subject's time-varying neural RDM, yielding a correlation value for each time point for each layer. The subject-averaged correlation values were then tested against chance.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

212

213 **Statistical inference**

214 In this study, we used Bayes factors (Dienes, 2011; Jeffreys, 1998; Rouder, Speckman, Sun, Morey, &
215 Iverson, 2009; Wagenmakers, 2007) to determine the evidence for the null and alternative hypotheses.
216 For the alternative hypothesis of above-chance decoding or correlation, a uniform prior was used
217 ranging from the maximum value observed during the baseline (before stimulus onset) up to 1 (e.g.,
218 100% decoding). For testing a non-zero difference between decoding accuracies, a uniform prior was
219 used ranging from the maximum absolute difference observed during the baseline up to 50% (0.5). We
220 then calculated the Bayes factor (BF) which is the probability of the data under the alternative
221 hypothesis relative to the null hypothesis. We thresholded BF>3 and BF>10 as substantial and strong
222 evidence for the alternative hypothesis, and BF<1/3 and BF<1/10 for substantial/strong evidence in
223 favour of the null hypothesis (Jeffreys, 1998; Wetzels et al., 2011). BF that lie between 1/3 and 3
224 indicate insufficient evidence for either hypothesis.

225

226 **Results**

227 We examined the representational dynamics of 200 different visual objects (Figure 1A), presented in
228 rapid and ultra-rapid sequences (Figure 1B) using EEG. During the sequences, participants detected
229 targets (boats or stars).

230

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

232 Participants were generally above chance (25%) at detecting targets (boats or stars) in the rapid and
233 ultra-rapid sequences (Figure 2A-B). There was no difference in performance between the boat and star
234 conditions (all BF < 1/3). On incorrect trials, responses often differed no more than one from the correct
235 answer (Figure 2, right columns). This indicates that in general, participants missed at most one target
236 when they responded incorrectly.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

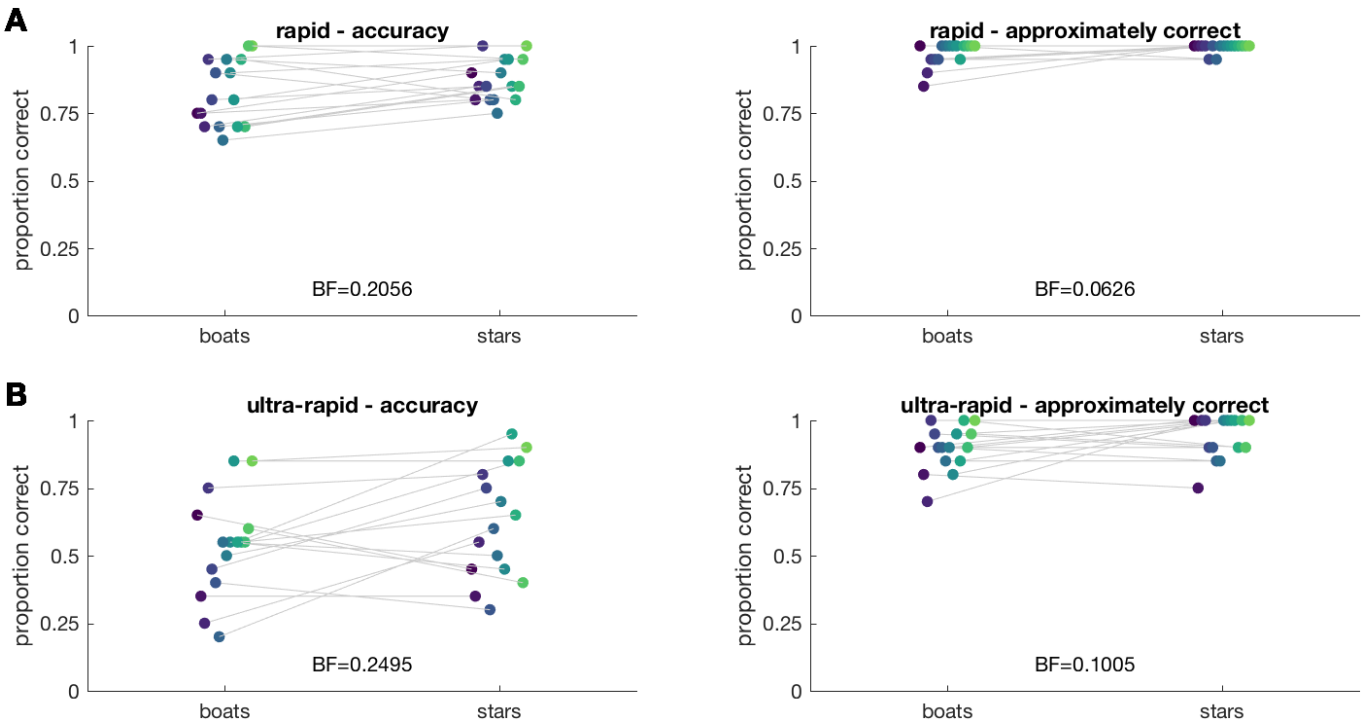


Figure 2: Behavioural results of target detection performance for rapid (A) and ultra-rapid (B) experimental sessions. Left columns show the mean proportion of correct responses for each participant separately for boat target sequences and star target sequences. Right columns show the mean approximately-correct (i.e., response differed by at most 1 from the correct answer) accuracy for each participant. Bayes Factors (BF) comparing accuracies on the boat and star sequences are listed above the x-axis.

The temporal dynamics of target selection were revealed by decoding targets from non-targets. The time-varying mean target-distractor decoding accuracy was computed separately for boat sequences and star sequences (Figure 3). Target-distractor decoding performance peaked around 75% in the rapid session (Figure 3A), and around 60% in the ultra-rapid session (Figure 3B). For both presentation durations, peak decoding performance was around 500ms. In the rapid session, we observed a second decoding peak around 1500ms. In both sessions, decoding for star targets was above chance earlier than for boats, and in the ultra-rapid session, decoding performance for star targets peaked higher than that for boats, which suggests that stars targets were easier to distinguish overall. Decoding performance remained above chance for 2 seconds for rapid sequences, and for 1 second in the ultra-rapid sequences.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

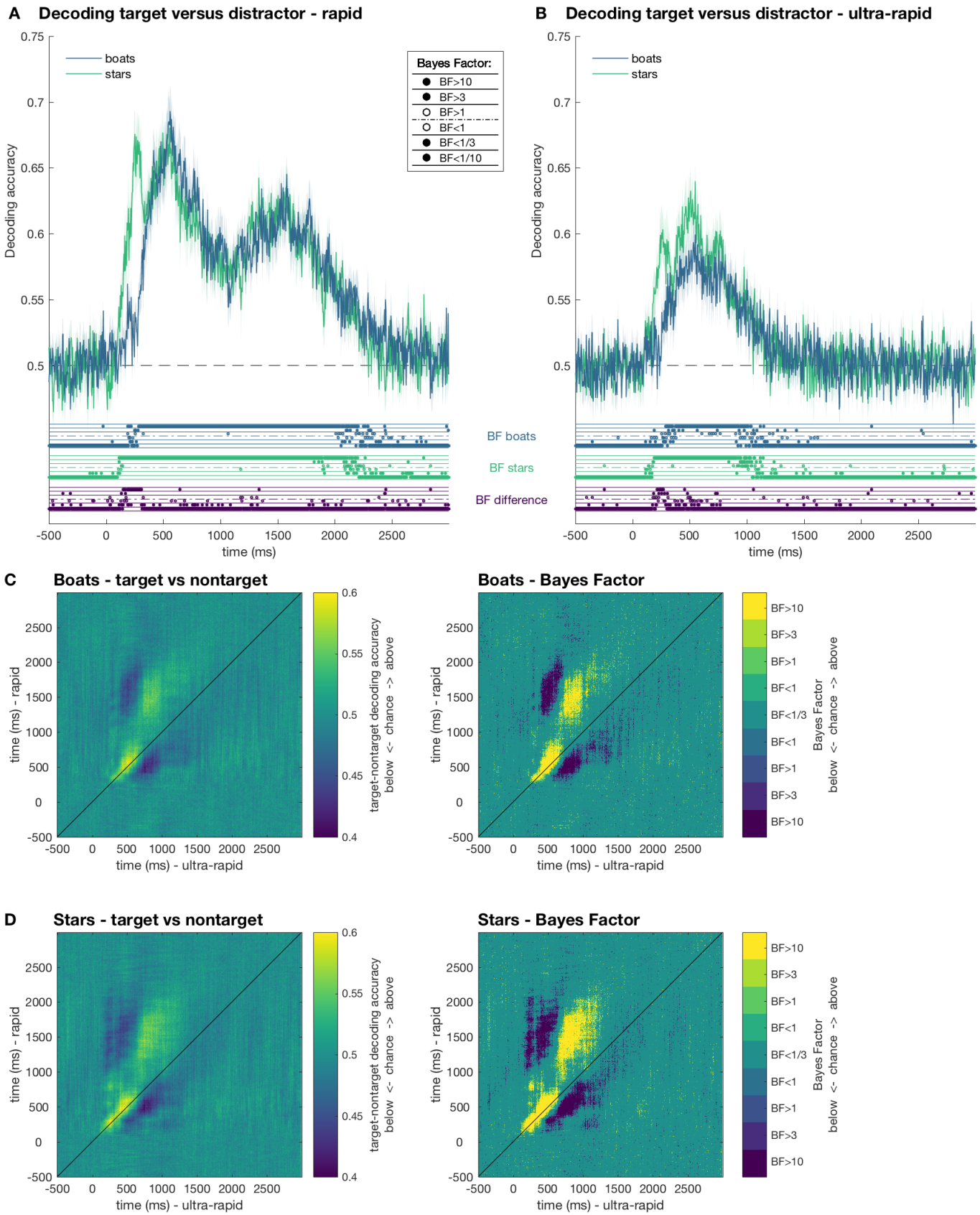


Figure 3. Decoding target versus distractor. For each target, the 4th, 5th, or 6th preceding distractor was selected, and classifiers were trained on target versus distractor. Plots show the mean leave-one-sequence-out cross-validated accuracy for the rapid 200ms duration session (A), and the ultra-rapid 50ms duration session (B). Results are shown separately for boat target sequences and star target

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sequences. Dots below plots indicate thresholded Bayes Factors (BF) for the boat (tow row) and star (middle row) sequences compared to chance and for the difference between boat and star sequences (bottom row). C-D temporal generalisation results. The left columns show classifier generalisation performance for the boat (C) and star (D) between the different presentation durations. The right columns show corresponding thresholded Bayes Factors (Yellow indicating above chance, and blue indicating below chance decoding). Higher than chance generalisation (yellow) above the diagonal indicates slower processing of images in the rapid session relative to the ultra-rapid session.

The temporal generalisation approach revealed differences between target selection in the rapid and ultra-rapid sequences. For both boat and star target sequences, the onset of target decoding was very similar between the rapid and ultra-rapid sequences, but diverged later in the time series (Figure 3c-d). In particular, late target-related processing was delayed in the rapid condition (~1200-1800ms) relative to the ultra-rapid condition (~750-1000ms), indicating that neuronal processes of target selection were actually faster for faster sequences.

Decoding categorical contrasts of 200 stimuli

In the rapid condition, we observed above chance decoding for all categorical levels (Figure 4, blue lines), starting at 100ms after stimulus onset. For the animacy level, the results showed two distinct peaks in decoding (200ms and 400ms). In contrast, peak decoding happened around 200ms for category and object decoding and 130ms for image decoding. For all categorical levels, above-chance decoding was sustained until around 600ms. Note that at 600ms, there were already two new stimuli presented.

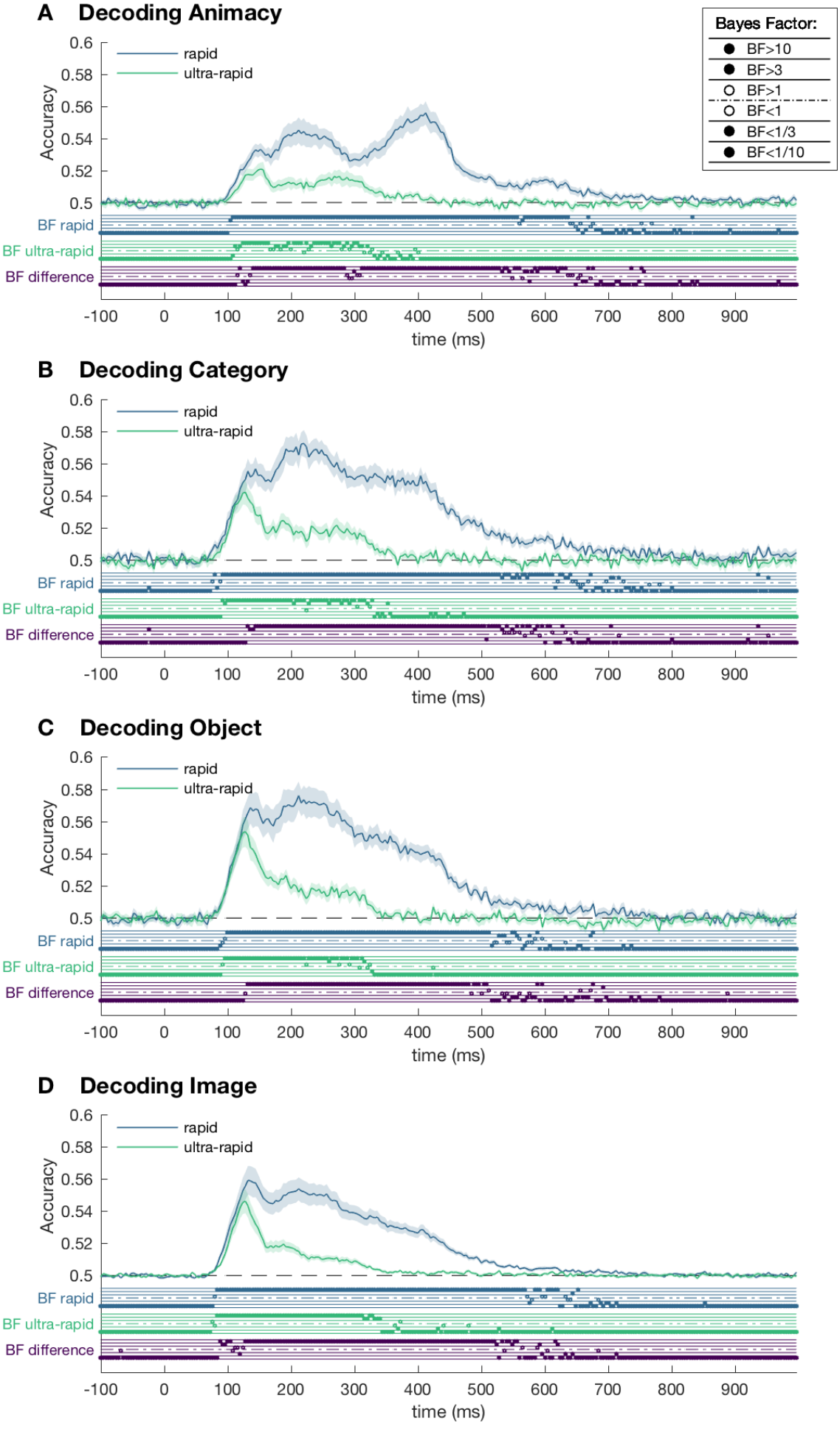
In the ultra-rapid condition (Figure 2, green lines), we again observed above-chance decoding for all levels. Notably, the onset of decoding was around the same time point as in the rapid session and subsequent decoding followed the same trajectory but diverged later in the time series (indicated by the bottom row of Bayes factors). The overall peak decoding performance was lower, and the peak decoding time points appeared earlier in the time series. Decoding remained above chance until around

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

300ms, which included five subsequent stimulus presentations. There was no difference between distractor processing on boat target and star target trials (BF < 1/10) for all categorical contrasts.

Temporal generalisation analyses were performed to compare categorical decoding between the rapid and ultra-rapid conditions. For all three categorical levels, we observed similar onsets between presentation durations, but longer subsequent processing for the rapid condition relative to the ultra-rapid condition (Figure 5). Notably, for the animacy distinction there was no evidence of generalisation between the rapid sequence around 500-600ms and the ultra-rapid sequence at any time point, despite a difference between decoding accuracies during this time period (as was seen in Figure 4). This suggests that a high-level animacy-related process was present in the slower condition but absent in the faster condition.

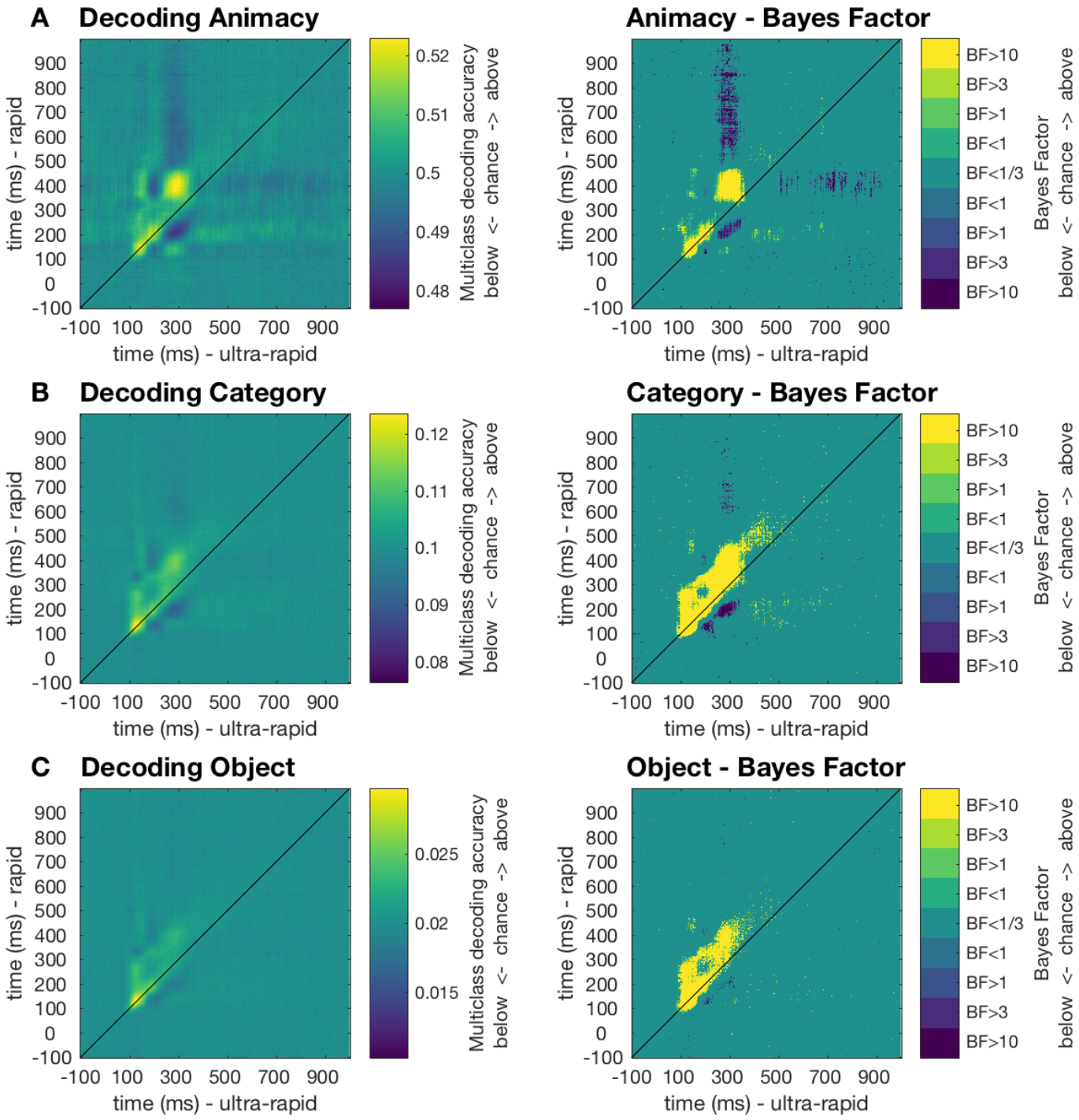
bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



300

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4. Mean decoding accuracy for rapid and ultra-rapid image duration sessions. A) Decoding animacy (animate versus inanimate). B) Mean pairwise decoding for the 10 categories (e.g., mammal, tools). C) Mean pairwise decoding for 50 object categories (e.g., dog, giraffe). D) Mean pairwise decoding for all 200 images. Shaded areas depict standard error of the mean across subjects. Dots below plots indicate thresholded Bayes Factors (BF) for the rapid session compared to chance (top rows), ultra-rapid session compared to chance (middle rows) and for the difference between rapid and ultra-rapid results (bottom rows).



bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5. Temporal generalisation results. The left columns show classifier generalisation performance for the three categorical levels (A-C) between the different presentation durations. The right columns show corresponding thresholded Bayes Factors (Yellow indicating above chance, and blue indicating below chance decoding). Higher than chance generalisation (yellow) above the diagonal indicates slower processing in the rapid session relative to the ultra-rapid session.

Representational dynamics of 200 stimuli

Emerging representational structures of the 200 stimuli were studied in the Representational Similarity Analysis (RSA) framework (Kriegeskorte & Kievit, 2013; Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008). A neural representational dissimilarity matrix (neural RDM) was created for each subject, and each time point containing the dissimilarities between all 200 stimuli. Neural RDMs were modelled as a linear combination of four candidate models; a low-level image silhouette model, and one model for each of the three categorical levels. We then analysed the mean beta estimates of the candidate models (Figure 6). For both presentation rates, the silhouette model captured the early response in the data, followed by the object, then category models. In the rapid condition, the animacy model emerged last, while in the ultra-rapid sequences the animacy model did not explain the neural RDM at any time point.

To visualise the dynamic representational structure, we created a two-dimensional embedding of all 200 images at three time points (200, 300, and 400ms), where the distance between images reflects their mean dissimilarity across subjects (Figure 7). The visualisations for the rapid session (Figure 7, left column) showed a high level animacy organisation. The visualisations for the ultra-rapid session (Figure 7, right column) did not show an animacy structure, but did show categorical clustering for humans, human faces and primates.

We further investigated the stage-wise computations by comparing the neural RDMs with RDMs created from deep neural network layer activation patterns (see e.g., Cichy, Khosla, Pantazis, Torralba, & Oliva,

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2016; Güçlü & Gerven, 2015; Khaligh-Razavi & Kriegeskorte, 2014). We compared the neural RDMs to RDMs obtained from the activation profiles of a pre-trained VGG-19 deep neural network (Simonyan & Zisserman, 2014). The deep neural network RDMs for each layer were correlated to the time-varying neural RDMs, yielding a correlation value for each time point for each layer. We found above chance correlations between the representations in the deep neural network layers, and the EEG responses (Figure 8). The correlations appeared later in the time series for later layers of the deep neural network.

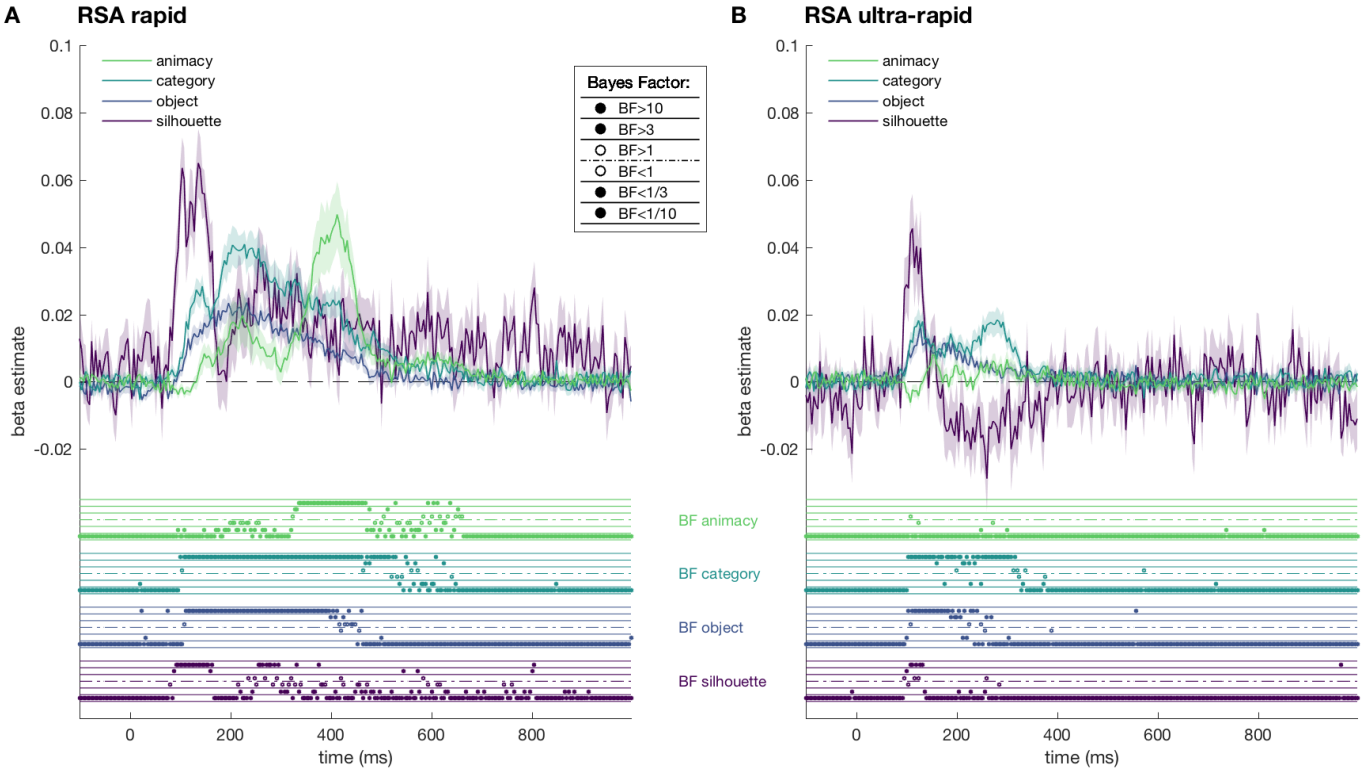
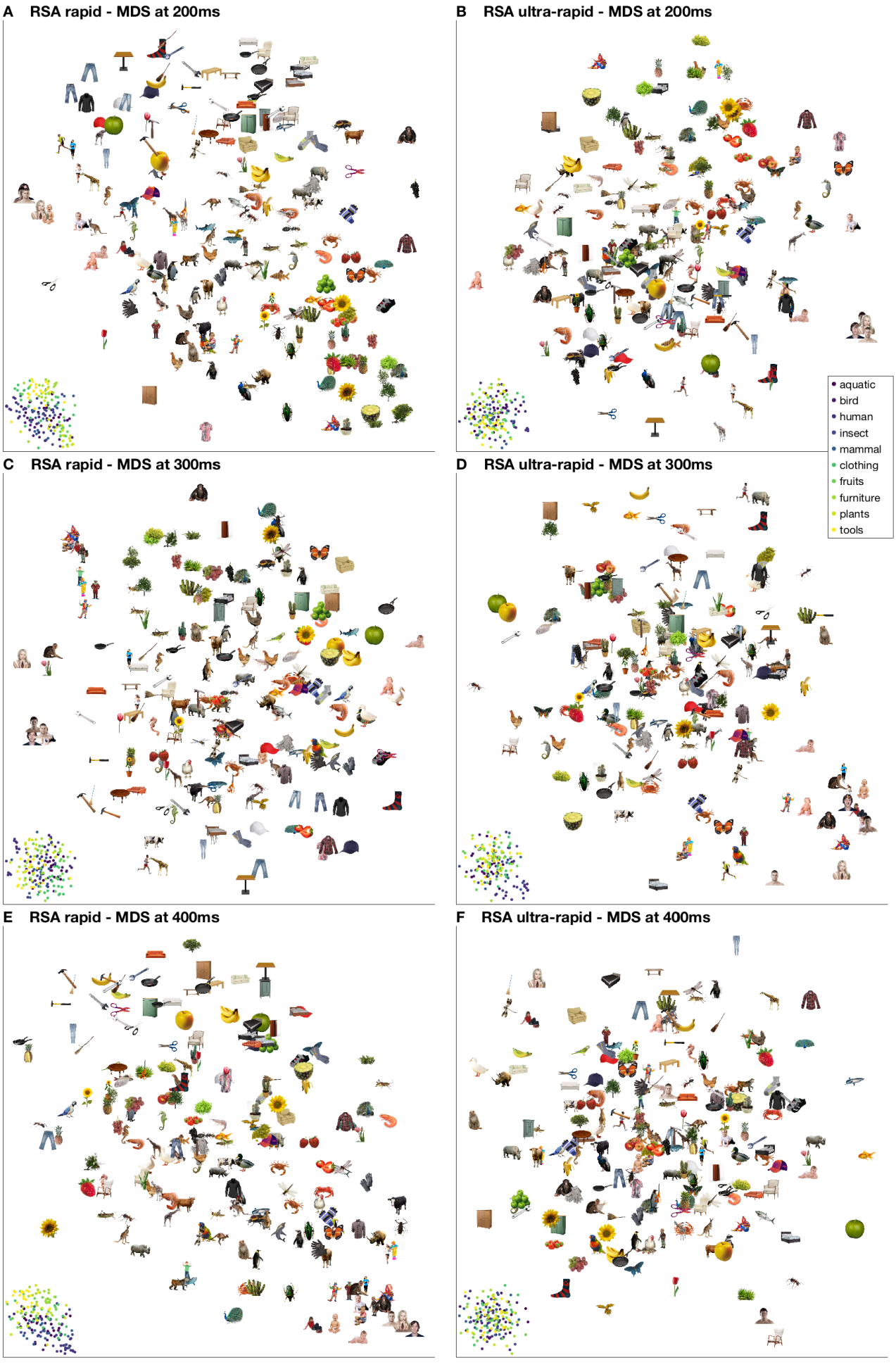


Figure 6. RSA model tests. The neural RDMs of each subject were modelled as linear combination of four candidate models, separately for the rapid sequences (A), and for the ultra-rapid sequences (B). Lines show estimated betas for the four models. Shaded areas reflect the standard error across subjects. Dots below plots indicate the thresholded Bayes Factors (BF) for each beta estimate.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



352

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 7. Representational structure at three time points. Stimuli are embedded in a two-dimensional space that reflects their pairwise distances. Stimuli that are shown further apart in this representation evoked more dissimilar neural responses. The rows reflect three time points in the response, and the left and right columns show data for the rapid and ultra-rapid conditions respectively. In the bottom left corner of each image, the same arrangement is shown, with images represented by dots coloured according to the 10 categories (see inset).

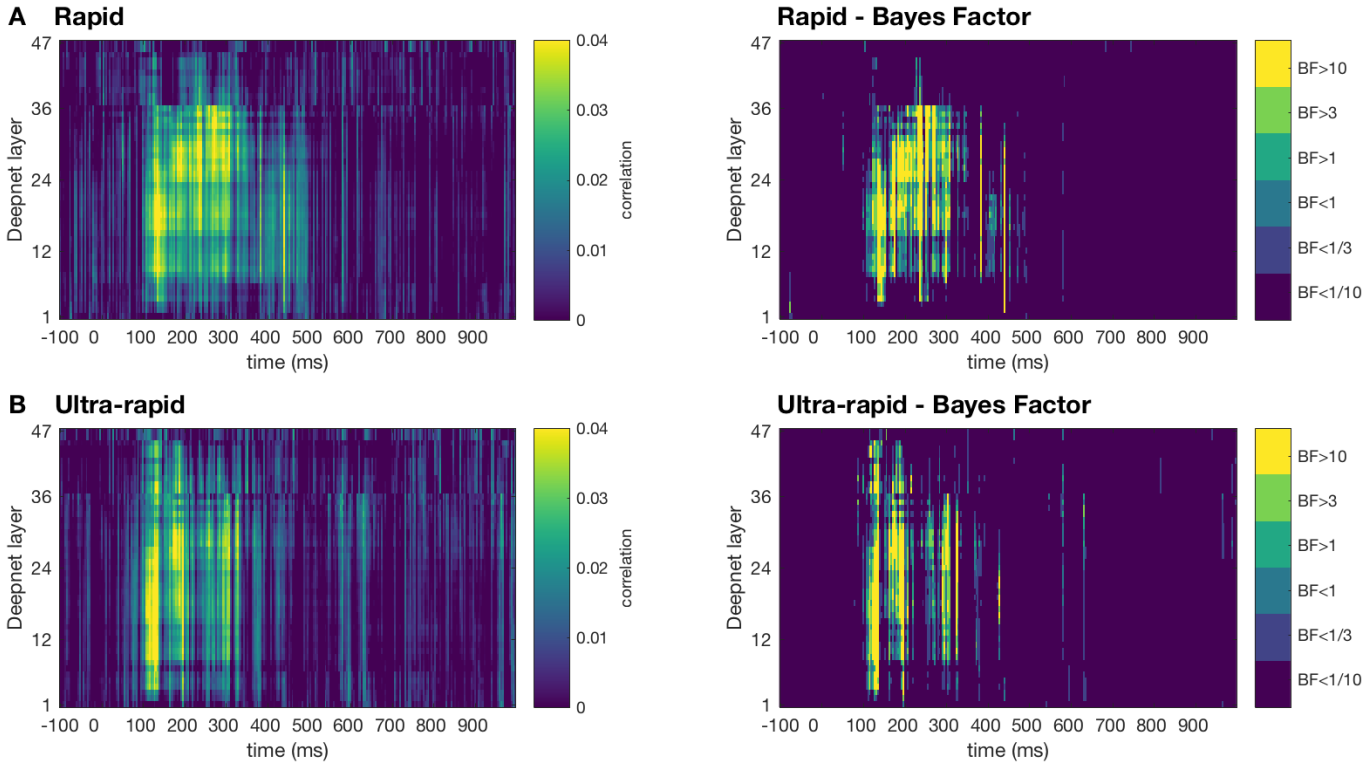


Figure 8. Correlation between neural RDMs and deep neural network layer activations. Subjects neural RDMs were correlated to RDMs constructed from deep neural network (VGG-19) activation patterns. The left columns show the mean correlations across subjects, and the right columns show the corresponding thresholded Bayes factors.

Discussion

In the current study, we characterised the representational dynamics of a large number of images in fast presentation sequences. Previous work has used MEG to investigate representations of much smaller image sets using slow image presentation paradigms; here we extend this work by looking at the representations of 200 objects during RSVP using standard 64-channel EEG. For rapid and ultra-rapid sequences, all 200 images could be decoded at four different categorical levels. Furthermore, neural responses to targets were distinct from those to distractor stimuli. Above-chance decoding outlasted

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

subsequent image presentations, supporting the idea that multiple object representations can co-exist in the visual system at different stages of processing (Marti & Dehaene, 2017). In keeping with the known hierarchical nature of the visual system, RSA model testing suggested neural responses relied on low-level visual features early in the time series, and subsequent processing was associated with increasing category abstraction (Carlson et al., 2013; Cichy et al., 2014). Correlations between EEG data and deep neural network activations revealed similarities between 100-300ms from image onset, suggesting that deep neural networks approximate early and mid-level visual responses. Overall, we show the unprecedented ability of the human brain to process images when pushing the limits of temporal perception.

Target decoding results revealed that neural responses to distractors diverged from star target responses much earlier than boat targets. This supports our initial hypothesis that star targets would be distinct from other images based on low-level visual features, unlike boat targets. The behavioural results, however, revealed target detection did not differ across boat and star trials, indicating that there was no “pop-out” effect of stars. This is despite anecdotal reports that participants found the star targets easier. Target versus distractor decoding for boats and stars peaked at 500ms, supporting previous evidence that high level cognitive processes mediate temporal selection (Marti & Dehaene, 2017; Sergent, Baillet, & Dehaene, 2005). These results suggest that distinguishable low-level features do not help with target detection in RSVP sequences, at least in the current design with such high variation in distractor images.

Target processing differed across the different experimental durations. In both the rapid and ultra-rapid sessions, targets could be distinguished from distractors for a long period of time, but this was exaggerated for the rapid session, where decoding was above chance for over 2 seconds, compared to 1 second in the ultra-rapid condition. Notably, target processing was much more prolonged than

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

categorical decoding for distractors, again indicative of higher level cognitive processes at play for target detection. Interestingly, however, the temporal dynamics of target responses did not appear to differ between the rapid and ultra-rapid sessions in terms of initial processing, but temporal generalisation showed that higher level processing stages occurred earlier for the ultra-fast sequences. Note that the current experimental design did not allow us to see which targets in the stream were missed, but effects are likely to be amplified for correctly detected targets. Indeed, Marti & Dehaene (2017) found that late responses were sustained for reported stimuli. Taken together, our results show that late target-related responses can speed up dramatically in faster sequences relative to slower sequences.

Neural responses to the 200 non-task-relevant (distractor) objects are indicative of fairly automatic early visual processing and divergence at later processing stages according to image duration. For all contrasts, image presentation duration and cognitive task set did not influence the earliest processing stages. When looking at decoding for the durations separately, onsets seemed to be earlier for the rapid than ultra-rapid conditions, in accordance with recent work showing earlier onsets for longer image durations (Mohsenzadeh et al., 2018). It is important to note, however, that higher signal strengths can also lead to earlier decoding onsets (Grootswagers, Wardle, et al., 2017), thus differences between onsets must be interpreted with caution in the context of larger peak decoding. Crucially, here Bayes factors revealed evidence that there was no difference in decoding at these early time points between the rapid and ultra-rapid image sequences (<150ms from image onset). Results from the temporal generalisation approach supported this view, by showing that initial processing stages occurred at the same time for the rapid and ultra-rapid sequences, as seen by the above-chance decoding on the diagonal in Figure 5. Finally, for the three categorical levels (animacy, category and object), Bayesian analyses revealed distractor processing did not differ between boat and star trials. These results suggest initial neural responses to all visual stimuli were similar regardless of their presentation duration.

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Despite similar early processing stages, later processing diverged according to image presentation duration. Representations during rapid sequences were stronger and lasted longer than those during ultra-rapid sequences, and temporal generalisation analyses showed that processes were prolonged for the rapid relative to the ultra-rapid condition. This echoes the target selection results showing that late processes occurred earlier for the faster sequences. Longer image durations seem to allow more consolidation, potentially due to recurrent processing. It is possible that longer durations allow time to reach some kind of threshold, which triggers further processing. Note that image duration and ISI are conflated in this design, so we cannot conclude whether or if stronger and longer processing occurs due to longer image presentation or due to delayed masking from the next stimulus. Future work can build on this approach to investigate the temporal limits of visual perception.

The RSA regression analyses provided insight into the differences in processing between the rapid and ultra-rapid sequences. The deep neural network decoding results (Figure 1) clearly showed that animacy, category and object can be decoded based on low-level visual features, demonstrating that the different category levels are inherently inter-related. The regression RSA technique attempts to dissociate the unique contributions of each model. In accordance with the decoding results, processes early in the time series (~100-150ms) were mostly explained by the low-level silhouette model for the rapid and ultra-rapid conditions (Carlson et al., 2011). Subsequent processing, however, elucidated the differential contributions of the different categorical contrasts, and how this varied for the different image durations. For the rapid condition, the category model appeared to have the largest unique contribution around 200ms, and the animacy model accounted for the most variance at about 400ms, indicating that increasing category abstraction occurred at higher levels of visual processing (Carlson et al., 2013; Cichy et al., 2014; Contini et al., 2017; Kriegeskorte, Mur, Ruff, et al., 2008). In contrast, the animacy model had no unique contribution to the signal for the in ultra-rapid sequences. The time course of the animacy model regression for the rapid condition (>350ms) suggests that the animate-

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

448 inanimate difference might exclusively account for the prolonged decoding in the rapid condition
449 relative to the ultra-rapid condition. This could imply that a high-level animacy effect requires sufficient
450 evidence accumulation to proceed, which does not happen at ultra-rapid presentation rates. The finding
451 that longer image presentations allow higher level processing is supported by steady-state visual evoked
452 potential (SSVEP) work showing that images presented at faster frequencies are biased towards earlier
453 visual processes in contrast to slower frequencies which allow higher level processing (Collins, Robinson,
454 & Behrmann, 2018). The other interesting thing to note about the rapid condition is that the low-level
455 silhouette model was also evident approximately 250-300ms after image onset, which was not present
456 in the ultra-rapid condition. This low-level re-emergence could reflect the response to the offset of the
457 image (Carlson et al., 2011).
458
459 In the visualisation of the representational structure, we observed a clear animate versus inanimate
460 organisation in the rapid presentation condition. At 200ms in the response, the structure reflected
461 mostly natural versus artificial, with plants, fruits and animals all clustering on one side. The structure at
462 400ms showed a clear animate – inanimate distinction (Caramazza & Shelton, 1998), which is commonly
463 observed in neural responses in the ventral temporal cortex (Cichy et al., 2014; Konkle & Caramazza,
464 2013; Kriegeskorte, Mur, Ruff, et al., 2008; Proklova, Kaiser, & Peelen, 2016) and has been shown to
465 match human categorisation behaviour well (Bracci & Op de Beeck, 2016; Carlson, Ritchie, Kriegeskorte,
466 Durvasula, & Ma, 2014; Grootswagers, Cichy, & Carlson, 2018; Mur et al., 2013; Ritchie, Tovar, &
467 Carlson, 2015). In the animate – inanimate organisation primates were located at the far end of the
468 animate side, which may reflect a continuum of biological classes in the brain (Connolly et al., 2012; Sha
469 et al., 2015) or typicality (Grootswagers, Ritchie, et al., 2017; Iordan, Greene, Beck, & Fei-Fei, 2016;
470 Posner & Keele, 1968; E. H. Rosch, 1973; E. Rosch & Mervis, 1975). No animacy structure was apparent
471 for the ultra-rapid condition, but rather individual categorical clusters emerged, such as human faces,
472 and, later, humans and primates as one category. Interestingly, in these visualisations gloves were

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

grouped with humans and primates, which could mean they were perceived as body parts, rather than inanimate objects. Our RSA results highlight the level of detail in the representation structure that can be obtained using EEG decoding and fast presentation rates. Here, we used a common 64-channel EEG, but future work can use this approach in combination with high-density EEG or other neuroimaging methods that are sensitive to finer spatial patterns, such as MEG.

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

Acknowledgements

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

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

497 **References**

498

499 Bankson, B. B., Hebart, M. N., Groen, I. I. A., & Baker, C. I. (2018). The temporal evolution of conceptual
500 object representations revealed through models of behavior, semantics and deep neural
501 networks. *NeuroImage*, *178*, 172–182. https://doi.org/10.1016/j.neuroimage.2018.05.037

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

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

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

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

514 Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction Time for Object
515 Categorization Is Predicted by Representational Distance. *Journal of Cognitive Neuroscience*,
516 *26*(1), 132–142. https://doi.org/10.1162/jocn_a_00476

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

519 Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural
520 networks to spatio-temporal cortical dynamics of human visual object recognition reveals
521 hierarchical correspondence. *Scientific Reports*, *6*. https://doi.org/10.1038/srep27755

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

522 Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time.
523 *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635
524 Collins, E., Robinson, A. K., & Behrmann, M. (2018). Distinct neural processes for the perception of
525 familiar versus unfamiliar faces along the visual hierarchy revealed by EEG. *NeuroImage*, *181*,
526 120–131. https://doi.org/10.1016/j.neuroimage.2018.06.080
527 Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., … Haxby, J. V. (2012).
528 The Representation of Biological Classes in the Human Brain. *The Journal of Neuroscience*, *32*(8),
529 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012
530 Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in
531 the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176.
532 https://doi.org/10.1016/j.neuropsychologia.2017.02.013
533 Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100
534 ms. *Journal of Vision*, *10*(4), 16. https://doi.org/10.1167/10.4.16
535 Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG
536 dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1),
537 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009
538 Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on*
539 *Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920
540 Greene, M. R., & Hansen, B. C. (2018). Shared spatiotemporal category representations in biological and
541 artificial deep neural networks. *PLOS Computational Biology*, *14*(7), e1006327.
542 https://doi.org/10.1371/journal.pcbi.1006327
543 Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and
544 its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548.
545 https://doi.org/10.1038/nrn3747

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

546 Grootswagers, T., Cichy, R. M., & Carlson, T. A. (2018). Finding decodable information that can be read
547 out in behaviour. *NeuroImage*, *179*, 252–262.
548 https://doi.org/10.1016/j.neuroimage.2018.06.022
549 Grootswagers, T., Ritchie, J. B., Wardle, S. G., Heathcote, A., & Carlson, T. A. (2017). Asymmetric
550 Compression of Representational Space for Object Animacy Categorization under Degraded
551 Viewing Conditions. *Journal of Cognitive Neuroscience*, *29*(12), 1995–2010.
552 https://doi.org/10.1162/jocn_a_01177
553 Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked
554 Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging
555 Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068
556 Güçlü, U., & Gerven, M. A. J. van. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of
557 Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–
558 10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015
559 Haynes, J.-D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and
560 Perspectives. *Neuron*, *87*(2), 257–270. https://doi.org/10.1016/j.neuron.2015.05.025
561 Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2016). Typicality sharpens category
562 representations in object-selective cortex. *NeuroImage*, *134*, 170–179.
563 https://doi.org/10.1016/j.neuroimage.2016.04.012
564 Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in
565 the human visual system. *Journal of Neurophysiology*, *111*(1), 91–102.
566 https://doi.org/10.1152/jn.00394.2013
567 Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.
568 Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed
569 by MEG and fMRI decoding. *Journal of Neurophysiology*, *115*(4), 2246–2250.
570 https://doi.org/10.1152/jn.01074.2015

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

571 Kaneshiro, B., Guimaraes, M. P., Kim, H.-S., Norcia, A. M., & Suppes, P. (2015). A Representational
572 Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification.
573 *PLOS ONE*, *10*(8), e0135697. https://doi.org/10.1371/journal.pone.0135697
574 Keysers, C., Xiao, D.-K., Földiák, P., & Perrett, D. I. (2001). The Speed of Sight. *Journal of Cognitive*
575 *Neuroscience*, *13*(1), 90–101. https://doi.org/10.1162/089892901564199
576 Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May
577 Explain IT Cortical Representation. *PLOS Comput Biol*, *10*(11), e1003915.
578 https://doi.org/10.1371/journal.pcbi.1003915
579 Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object Category Structure in Response Patterns of
580 Neuronal Population in Monkey Inferior Temporal Cortex. *Journal of Neurophysiology*, *97*(6),
581 4296–4309. https://doi.org/10.1152/jn.00024.2007
582 King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal
583 generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210.
584 https://doi.org/10.1016/j.tics.2014.01.002
585 Konkle, T., & Caramazza, A. (2013). Tripartite Organization of the Ventral Stream by Animacy and Object
586 Size. *Journal of Neuroscience*, *33*(25), 10235–10242. https://doi.org/10.1523/JNEUROSCI.0983-
587 13.2013
588 Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation,
589 and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412.
590 https://doi.org/10.1016/j.tics.2013.06.007
591 Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational Similarity Analysis - Connecting
592 the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.
593 https://doi.org/10.3389/neuro.06.004.2008

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(Nov), 2579–2605.

Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? *European Journal of Neuroscience*, *21*(7), 2007–2018. https://doi.org/10.1111/j.1460-9568.2005.04029.x

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*(1), 28–35. https://doi.org/10.3758/PBR.15.1.28

Mack, M. L., & Palmeri, T. J. (2011). The Timing of Visual Object Categorization. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00165

Mack, M. L., & Palmeri, T. J. (2015). The Dynamics of Categorization: Unraveling Rapid Categorization. *Journal of Experimental Psychology: General*, No Pagination Specified. https://doi.org/10.1037/a0039184

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences*, *15*(3), 97–103. https://doi.org/10.1016/j.tics.2011.01.004

Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nature Communications*, *8*(1), 1955. https://doi.org/10.1038/s41467-017-02079-x

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex. *Journal of Neurophysiology*, *100*(3), 1407–1419. https://doi.org/10.1152/jn.90248.2008

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

618 Mohsenzadeh, Y., Qin, S., Cichy, R., & Pantazis, D. (2018). Ultra-Rapid Serial Visual Presentation Reveals
619 Dynamics of Feedforward and Feedback Processes in the Ventral Visual Pathway. *BioRxiv*,
620 350421. https://doi.org/10.1101/350421
621 Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human Object-
622 Similarity Judgments Reflect and Transcend the Primate-IT Object Representation. *Frontiers in*
623 *Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00128
624 Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and
625 ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. https://doi.org/10.1016/S1388-
626 2457(00)00527-7
627 Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern
628 Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*.
629 https://doi.org/10.3389/fninf.2016.00027
630 Peelen, M. V., & Caramazza, A. (2012). Conceptual Object Representations in Human Anterior Temporal
631 Cortex. *Journal of Neuroscience*, *32*(45), 15728–15736.
632 https://doi.org/10.1523/JNEUROSCI.1953-12.2012
633 Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental*
634 *Psychology*, *77*(3p1), 353. http://dx.doi.org/10.1037/h0025953
635 Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965–966.
636 Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology:*
637 *Human Learning and Memory*, *2*(5), 509.
638 Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms
639 per picture. *Attention, Perception, & Psychophysics*, *76*(2), 270–279.
640 Proklova, D., Kaiser, D., & Peelen, M. (2017). MEG sensor patterns reflect perceptual but not categorical
641 similarity of animate and inanimate objects. *BioRxiv*, 238584. https://doi.org/10.1101/238584

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

642 Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling Representations of Object Shape and
643 Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. *Journal of*
644 *Cognitive Neuroscience*, 1–13. https://doi.org/10.1162/jocn_a_00924
645 Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging Object Representations in the Visual System
646 Predict Reaction Times for Categorization. *PLoS Comput Biol*, *11*(6), e1004316.
647 https://doi.org/10.1371/journal.pcbi.1004316
648 Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350.
649 https://doi.org/10.1016/0010-0285(73)90017-0
650 Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories.
651 *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9
652 Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images
653 reveals a robust face-selective electrophysiological response in the human brain. *Journal of*
654 *Vision*, *15*(1), 18–18. https://doi.org/10.1167/15.1.18
655 Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting
656 and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.
657 Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization
658 of natural images. *Nature Neuroscience*, *5*(7), 629–630. https://doi.org/10.1038/nn866
659 Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to
660 consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391.
661 Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C.
662 (2015). The Animacy Continuum in the Human Ventral Vision Pathway. *Journal of Cognitive*
663 *Neuroscience*, *27*(4), 665–678. https://doi.org/10.1162/jocn_a_00733
664 Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image
665 Recognition. In *International Conference on Learning Representations*. Retrieved from
666 http://arxiv.org/abs/1409.1556

bioRxiv preprint doi: https://doi.org/10.1101/394148; this version posted August 17, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

667 Teichmann, L., Grootswagers, T., Carlson, T., & Rich, A. N. (2018). Decoding Digits and Dice with
668 Magnetoencephalography: Evidence for a Shared Representation of Magnitude. *Journal of*
669 *Cognitive Neuroscience*, *30*(7), 999–1010. https://doi.org/10.1162/jocn_a_01257
670 Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*,
671 *381*(6582), 520–522. https://doi.org/doi:10.1038/381520a0
672 VanRullen, R., & Thorpe, S. J. (2001). The Time Course of Visual Processing: From Early Perception to
673 Decision-Making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461.
674 https://doi.org/10.1162/08989290152001880
675 Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic*
676 *Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105
677 Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S.-M., & Carlson, T. A. (2016).
678 Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured
679 with MEG. *NeuroImage*, *132*, 59–70. https://doi.org/10.1016/j.neuroimage.2016.02.019
680 Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical
681 Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives*
682 *on Psychological Science*, *6*(3), 291–298. https://doi.org/10.1177/1745691611406923
683
684