

1 **Typical real-world locations impact the time course of object coding**

2

3 Daniel Kaiser<sup>1,\*</sup>, Merle M. Moeskops<sup>1</sup>, Radoslaw M. Cichy<sup>1</sup>

4

5 <sup>1</sup>*Department of Education and Psychology, Freie Universität Berlin, 14195 Berlin, Germany*

6

7 \*Correspondence to:

8 Dr. Daniel Kaiser

9 Department of Education and Psychology

10 Freie Universität Berlin

11 Habelschwerdter Allee 45

12 14195 Berlin, Germany

13 [danielkaiser.net@gmail.com](mailto:danielkaiser.net@gmail.com)

14 **Abstract**

15 Everyday visual environments are spatially structured in that objects often appear at  
16 typical locations in space: for example, lamps hang from the ceiling, whereas carpets lie  
17 on the floor. As a consequence, objects repeatedly occupy similar visual field locations.  
18 The long-term experience with these spatial regularities prompts the hypothesis that the  
19 visual system is tuned to such retinotopic object locations. A key prediction is that  
20 typically positioned objects should be coded more efficiently. To test this prediction, we  
21 recorded electroencephalography (EEG) while participants viewed briefly presented  
22 objects appearing in their typical locations (e.g., an airplane in the upper visual field) or in  
23 atypical locations (e.g., an airplane in the lower visual field). Multivariate pattern analysis  
24 applied to the EEG data revealed that object classification depended on positional  
25 regularities: Objects were classified more accurately when positioned typically, rather  
26 than atypically, already at 140 ms, suggesting that relatively early stages of object  
27 processing are tuned to typical retinotopic locations. Our results confirm the prediction  
28 that long-term experience with objects occurring at specific locations leads to enhanced  
29 perceptual processing when these objects appear in their typical locations. This may  
30 indicate a neural mechanism for efficient natural scene processing, where a large number  
31 of typically positioned objects needs to be processed.

## 32 **1 Introduction**

33 Visual objects are enclosed entities that can in principle be moved around freely.  
34 However, in everyday environments object positions are often quite constrained. For  
35 instance, consider the predictability in the locations of objects in a living room: The sofa is  
36 facing the TV, a table is in between the two, a lamp hangs from the ceiling, whereas  
37 carpets lie on the floor. This example illustrates that the object content of natural scenes  
38 is organized in repeatedly occurring positional structures (Bar, 2000; Chun, 2002). Many  
39 previous studies have investigated how inter-object relationships in these positional  
40 structures (e.g., lamps appearing above tables) impact behavioral performance and  
41 neural processing (Biederman, Mezzanotte, & Rabinowitz, 1982; Kaiser, Stein, & Peelen,  
42 2014; Oliva & Torralba, 2007; Wolfe, Võ, Evans, & Greene, 2011). However, positional  
43 object structures often also imply that individual objects are associated with particular  
44 locations in space (e.g., lamps appearing in the upper part of a scene). It has recently  
45 been proposed that the visual system is tuned to these regularities (Kaiser & Haselhuhn,  
46 2017; Kravitz, Vinson, & Baker, 2008), which could facilitate neural processing for  
47 individual objects appearing in their typical real-world locations.

48         Such location-specific variations in object coding are suggested by previous results  
49 that indicate the co-representation of object identity and location information in visual  
50 cortex: (1) cortical responses depend on the position of the object in the visual field  
51 (Hemond, Kanwisher, & Op de Beeck, 2007; Hasson, Levy, Behrmann, Hendler, & Malach,  
52 2002), (2) object selective cortex contains information about both an object's identity and  
53 its location (Cichy, Chen, & Haynes, 2011; Golomb & Kanwisher, 2011; Hong, Yamins, Majaj,  
54 & DiCarlo, 2017; Kravitz, Kriegeskorte, & Baker, 2010; Schwarzlose, Swisher, Dang, &  
55 Kanwisher, 2008), and (3) information about object identity and location emerge at

56 similar time points in visual processing (Isik, Meyers, Leibo, & Poggio, 2014; Carlson,  
57 Hogendoorn, Kanai, Mesik, & Turret, 2011).

58         The link between identity and location information in object processing creates  
59 the possibility that the two properties interact. In everyday environments, the visual  
60 system is repeatedly faced with positional structures, where individual object positions  
61 are highly predictable. Through this repeated exposure, retinotopic object-coding  
62 mechanisms could get tuned to typical object locations, forming neural channels that  
63 integratively process an object's identity and its location. Such location-specific  
64 processing channels would enhance the processing of an object when it appears in its  
65 typical locations within a scene – and within the visual field. Evidence for such a  
66 processing enhancement has been found in the domain of person perception, where  
67 typical configurations impact cortical responses to individual face and body parts (Chan,  
68 Kravitz, Truong, Arizpe, & Baker, 2010; de Haas et al., 2016; Henriksson, Mur, &  
69 Kriegeskorte, 2015). For example, in face-selective visual cortex, response patterns are  
70 better discriminable for typically, as compared to atypically, positioned face parts (de  
71 Haas et al., 2016), revealing visual processing channels that are tuned to the spatial  
72 regularities in the face.

73         Here, we test the prediction that the positional regularities contained in natural  
74 scenes can similarly facilitate the processing of everyday objects appearing in their typical  
75 locations. Participants viewed objects associated with upper and lower visual field  
76 locations (e.g., a lamp or a carpet) (Figure 1A) while we recorded electroencephalography  
77 (EEG). We used multivariate classification on the EEG data (Contini, Wardle, & Carlson,  
78 2017) to track the time course of object coding with high temporal precision. Analyses  
79 revealed that after 140ms visual processing of objects is affected by their typical real-

80 world locations: Objects appearing in their typical locations (e.g., a lamp in the upper  
81 visual field and a carpet in the lower visual field) could be decoded more successfully than  
82 objects appearing in atypical locations (e.g., a carpet in the upper visual field and a lamp  
83 in the lower visual field). These results suggest that early, rather than late, stages of  
84 visual processing are tuned to the positional object structure of real-world scenes.

## 85 **2 Materials and Methods**

### 86 *2.1 Participants*

87 Thirty-four healthy adults (mean age 26.4 years,  $SD = 5.4$ ; 23 female) completed  
88 the experiment. The sample size was set a-priori, based on considerations regarding  
89 statistical power: A sample size of 34 is needed for detecting a simple effect with a  
90 medium effect size of  $d = 0.5$  with a probability of more than 80%. All participants had  
91 normal or corrected-to-normal vision, provided informed consent and received monetary  
92 reimbursement or course credits for their participation. All procedures were approved by  
93 the ethical committee of the Department of Education and Psychology of the Freie  
94 Universität Berlin and were in accordance with the Declaration of Helsinki.

### 95 *2.2 Stimuli*

96 The stimulus set consisted of greyscale images of six objects associated with  
97 typical visual field locations, of which three were associated with upper visual field  
98 locations (lamp, airplane, and hat) and three were associated with lower visual field  
99 locations (carpet, boat, and shoe). For each object, ten exemplars were used (see Figure  
100 1A for stimulus examples). The images were matched for overall luminance (using the  
101 SHINE toolbox; Willenbockel et al., 2010) and displayed on a white background.

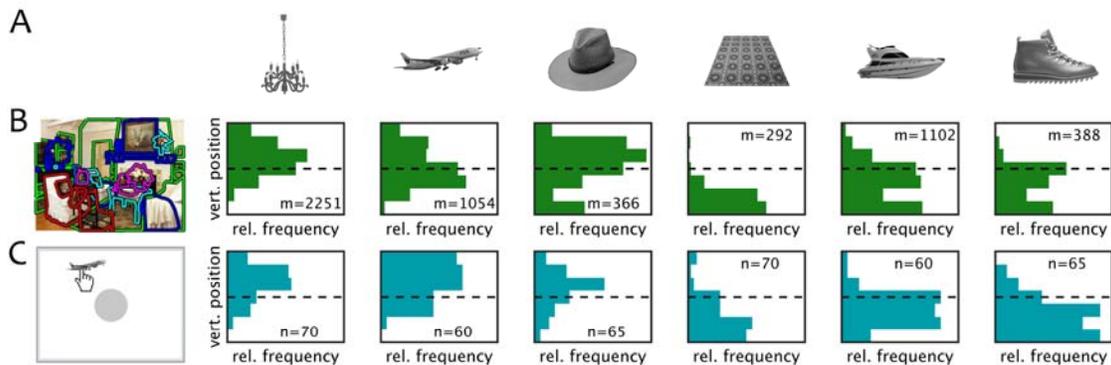
102 To ensure that the six objects could be reliably linked to a specific location, we  
103 validated the association of the six objects with a specific part of the visual field in two  
104 ways. First, we assessed the typical spatial distribution of each object in natural scenes,  
105 assuming that natural scene photographs represent a snapshot of the visual field roughly  
106 approximating natural viewing conditions. Hence, the distribution of the objects in the  
107 scene image should be similar to their distribution across the visual field. To objectively  
108 measure the typical position of each object within a scene, we queried a huge number of

109 scene photographs (>100,000) from the LabelMe toolbox, where human observers  
110 annotated single objects by drawing labelled polygons (Russell, Torralba, Murphy, &  
111 Freeman, 2008). For all scenes containing a specific object we computed the mean pixel  
112 coordinate of the area labeled as belonging to the object and then averaged these  
113 positions across scenes. The resulting “typical” object locations showed that, as  
114 expected, the upper visual field objects were associated with locations ( $y$ : vertical  
115 coordinate from bottom (0) to top (1) of the scene) in the upper parts of scenes (lamp:  $y$  =  
116 0.61,  $SD$  = 0.17; airplane:  $y$  = 0.52,  $SD$  = 0.20; hat:  $y$  = 0.53,  $SD$  = 0.25), while lower visual  
117 field objects were associated with locations in the lower part of scenes (carpet:  $y$  =  
118 0.17,  $SD$  = 0.13; boat:  $y$  = 0.33,  $SD$  = 0.23; shoe:  $y$  = 0.31,  $SD$  = 0.20). The typical location in  
119 scenes differed significantly between objects associated with the upper and lower visual  
120 field,  $t > 13.3$ ,  $p < .001$ , for all pairwise comparisons. Figure 1B shows the distribution of  
121 object locations along the vertical axis of the scenes, split into 7 bins.

122         Second, we sought to demonstrate a correspondence between this automated,  
123 scene-based measure and people’s explicit associations of the objects with particular  
124 locations in space. We thus asked a set of participants (between 60 and 70 participants;  
125 including the participants of the current study, after the completion of the EEG  
126 experiment) to indicate the typical locations in which they expect to see each of the six  
127 objects. In this task, participants were asked to drag the image of a single exemplar of  
128 each object to its typical location on a computer screen (imagining that the computer  
129 screen represents their field of view in a natural scene). The central part of the screen –  
130 where the object initially appeared – was blocked (indicated by a grey circle), so that  
131 participants (1) could not place the object in a central location of the screen, and (2) had  
132 to move the object before proceeding. As expected, participants more often chose upper

133 screen positions ( $y$ : vertical coordinate from bottom (0) to top (1) of the screen) for the  
134 upper visual field objects (lamp:  $y = 0.65$ ,  $SD = 0.19$ ; airplane:  $y = 0.67$ ,  $SD = 0.20$ ; hat:  $y =$   
135  $0.57$ ,  $SD = 0.20$ ), and lower screen positions for the lower visual field objects (carpet:  $y =$   
136  $0.29$ ,  $SD = 0.22$ ; boat:  $y = 0.36$ ,  $SD = 0.18$ ; shoe:  $y = 0.30$ ,  $SD = 0.18$ ). The vertical locations  
137 chosen by the participants differed significantly between objects associated with the  
138 upper and lower visual field,  $t > 6.04$ ,  $p < .001$ , for all pairwise comparisons. Figure 1C  
139 shows the distribution of vertical object locations on the screen, split into 7 bins. The  
140 scene-based measure and participants' explicit assessment thus provided converging  
141 evidence for the association of the objects with specific spatial locations.

142



143

144 **Figure 1.** A) Example Stimuli. Ten different exemplar images of six objects each (here, one  
145 example image for each object shown) were used as stimuli, of which three objects were  
146 associated with the upper visual field (lamp, airplane, hat) and three were associated  
147 with the lower visual field (carpet, boat, shoe). B) To validate our assessment of visual  
148 field associations, we automatically extracted the positions for each object in a large set  
149 of labelled scene photographs taken from the LabelMe scene database (Russell et al.,  
150 2008). For each scene, we determined the relative position of the object along the  
151 vertical axis, and plotted the distribution across 7 bins ( $m$ : number of scenes for each  
152 object). C) Additionally, we asked a group of participants to indicate for each object the

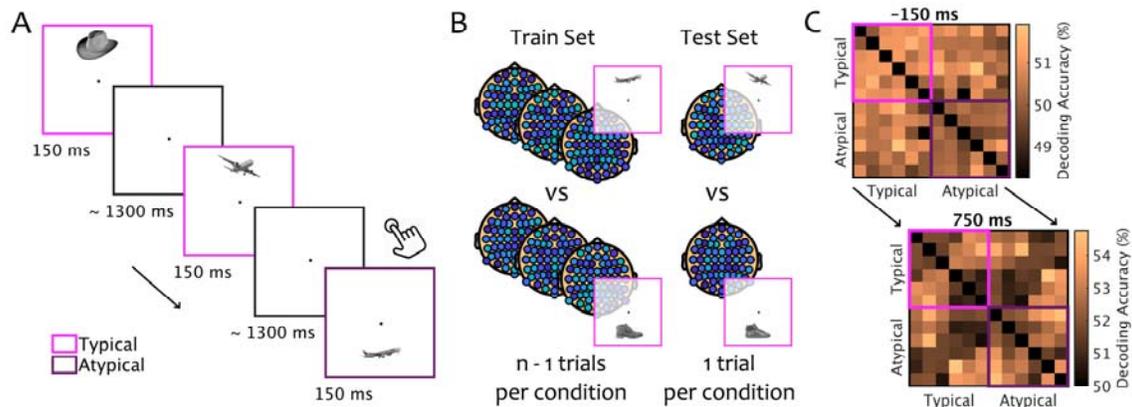
153 position that it typically occupies in the visual field by dragging the object to the desired  
154 location. We then computed the distribution of relative locations along the vertical axis of  
155 the screen, split into 7 bins ( $n$ : number of participants that indicated the typical location  
156 for each object). Both measures confirmed the spatial priors associated with the six  
157 objects.

158

### 159 2.3 Experimental Design

160 To test whether objects are processed differently when presented in typical and  
161 atypical locations within the visual field, the objects were presented in the upper or lower  
162 visual field (Figure 2A). On every trial, one object exemplar was presented in one of the  
163 two locations for 150ms, followed by a variable inter-trial interval (randomly jittered, from  
164 1250ms to 1750ms). Stimuli were presented at  $3.25^\circ$  vertical eccentricity and subtended a  
165 visual angle of maximally  $3^\circ$  in horizontal and vertical axes. Stimulus presentation was  
166 controlled using the Psychtoolbox (Brainard, 1997). Participants were asked to detect  
167 one-back repetitions on an object level (e.g., two different airplanes in direct succession;  
168 see Figure 2A). Repetitions occurred on 13% of the trials and equally often for typically and  
169 atypically positioned repetition targets and for the top and bottom locations. Participants  
170 performed accurately on this task (mean accuracy 96%,  $SD = 2\%$ ), with no difference in  
171 accuracy between typically and atypically positioned objects,  $t(33) = 0.87$ ,  $p = .391$ . One-  
172 back repetition trials were removed from all EEG analyses. The whole experiment  
173 consisted of 1656 trials (including 216 repetition trials). The 1440 non-repetition trials  
174 consisted of 12 repetitions of each object exemplar in each location (i.e., 120 repetitions  
175 per object and location). The experiment was split into 8 runs, and participants could take  
176 breaks between the runs. Twelve participants completed an extended experimental

177 session with 2760 trials (including 360 repetition trials), which additionally contained the  
178 same conditions at large eccentricities in half of the trials; these additional data are not  
179 reported here. The 2400 non-repetition trials consisted of 10 repetitions of each object  
180 exemplar in each of the four locations (i.e., 100 repetitions per object and location). The  
181 extended experiment was split into 12 runs.  
182



183  
184 **Figure 2.** Paradigm and Classification Logic. A) Stimuli were presented for 150 ms in upper  
185 or lower visual field locations, corresponding to an object's typical or an atypical location.  
186 Participants were instructed to detect occasional one-back repetitions on an object-  
187 category level (irrespective of the stimulus location) by pressing a button. Colors  
188 indicating the two regularity conditions are shown for illustrative purposes only. B)  
189 Multivariate classification was performed on response patterns across all electrodes,  
190 separately for each pairwise combination of objects (exemplified here by airplane and  
191 shoe in regular locations). The data was split into two sets: a training set consisting of all  
192 (but one) trials for each object and a testing set consisting of the two left-out trials. LDA  
193 classifiers were repeatedly trained and tested until every trial was left out once and  
194 accuracy was averaged across these repetitions. C) The pairwise classification analysis  
195 was repeated for each 10 ms time bin, resulting in a 12-by-12 matrix of pairwise

196 classification accuracies (with an empty diagonal) at every time point. To determine  
197 differences between typically and atypically positioned objects, pairwise comparisons  
198 within the typically placed objects (pink rectangle, upper left) and within the atypically  
199 placed objects (purple rectangle, lower right) were averaged and compared (Figure 3).

200

#### 201 *2.4 EEG recording and preprocessing*

202 The EEG was recorded using an EASYCAP 64-channel system and a Brainvision  
203 actiCHamp amplifier. The 64 electrodes were arranged in accordance with the standard  
204 10-10 system. The data was recorded at a sampling rate of 1000 Hz and filtered online  
205 between 0.5 and 70 Hz. For one participant, due to a technical problem, only data from 32  
206 electrodes was recorded. All electrodes were referenced online to the Fz electrode.

207 Offline preprocessing was performed in MATLAB, using the FieldTrip toolbox  
208 (Oostenveld, Fries, Maris, & Schoffelen, 2011). The continuous EEG data was epoched into  
209 trials ranging from 150ms before stimulus onset to 750ms after stimulus onset. Trials  
210 containing movement-related artefacts were visually identified and excluded from all  
211 analyses. Blink and eye movement artifacts were identified and removed using  
212 Independent Components Analysis (ICA) and visual inspection of the resulting  
213 components. To increase the signal-to-noise ratio of the classification analyses (Carlson,  
214 Tovar, Alink, & Kriegeskorte, 2013), the data was downsampled to 100Hz.

#### 215 *2.5 EEG classification procedure*

216 Multivariate classification analyses were carried out in MATLAB using the  
217 CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016). Classification was performed  
218 separately for each 10ms time bin, resulting in classification time courses with 10 ms  
219 resolution. The analysis was performed pairwise, for all possible combinations of the six

220 objects appearing in the two locations. Linear discriminant analysis (LDA) classifiers were  
221 always trained and tested on data from two conditions (e.g., an airplane in the upper  
222 visual field versus a carpet in the lower visual field), using a leave-one-out partitioning  
223 scheme (Figure 2B). The testing set consisted of all but one trials for each of the two  
224 conditions, while one trial for each of the two conditions was held back and used for  
225 classifier testing. This procedure was repeated until every trial was left out once. Classifier  
226 performance was averaged across these repetitions. The pairwise decoding analysis  
227 resulted in 12-by-12 matrix of decoding accuracies at each time point (reflecting all  
228 comparisons between the six objects appearing in the two locations) (Figure 2C).

## 229 *2.6 Overall classification dynamics*

230 To assess the overall classification dynamics over time, we computed the general  
231 discriminability of the twelve different conditions. All pairwise classification accuracies  
232 were averaged, revealing a time course of object decoding independently of the  
233 positional regularities. This time course of overall classification accuracy was used to  
234 define time points of interest at the peaks of the classification time series, where  
235 classification performance was particularly pronounced. Using a “region of interest” logic  
236 frequently applied in fMRI analyses (Poldrack, 2007), we used these peaks as “time points  
237 of interest” to increase the detection power of subsequent analyses.

## 238 *2.7 Object classification in typical and atypical locations*

239 To determine an effect of positional regularity on object decoding, we compared  
240 performance when classifying among typically positioned objects versus among atypically  
241 positioned objects. Pairwise classification accuracies were averaged for all comparisons  
242 between typically positioned objects (e.g., an airplane in the upper visual field versus a  
243 shoe in the lower visual field) and for all comparisons between atypically positioned

244 objects (e.g., a shoe in the upper visual field versus an airplane in the lower visual field)  
245 (Figure 2C). Subsequently, the two resulting classification time series (for typically and  
246 atypically positioned objects) were compared. To increase the statistical power of this  
247 comparison, we specifically focused on the effect of positional regularity at the peaks in  
248 overall classification.

#### 249 2.8 *Sensor-space searchlight analysis*

250 To investigate which sensors contributed most to the observed effects, we  
251 performed a sensor-space searchlight analysis. For this analysis, the pairwise classification  
252 procedure was repeated for neighborhoods of seven adjacent electrodes around each  
253 individual electrode; the resulting classification accuracy was then mapped onto a scalp  
254 representation. This procedure allowed us to infer the approximate spatial distribution of  
255 classification differences between typically and atypically positioned objects. As for one  
256 participant only data from 32 electrodes was available, this participant was not included  
257 in the searchlight analysis.

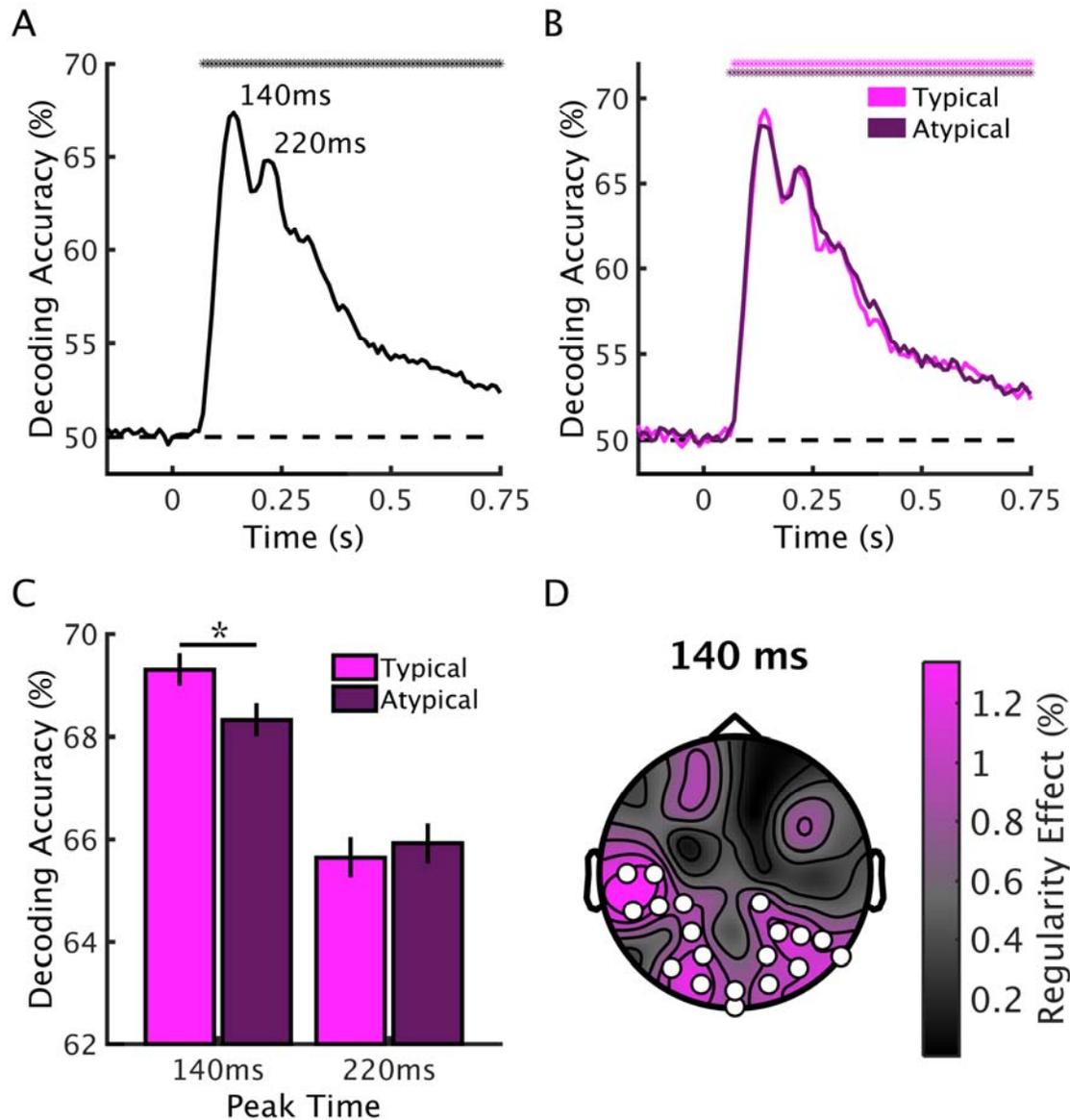
#### 258 2.9 *Statistical testing*

259 To identify significant effects across time we used a threshold-free cluster-  
260 enhancement procedure (Smith & Nichols, 2009) with default parameters. Multiple  
261 comparison correction was based on a sign-permutation test (with null distributions  
262 created from 10,000 bootstrapping iterations) as implemented in CoSMoMVA  
263 (Oosterhof et al., 2016). The resulting statistical maps were thresholded at  $Z > 1.96$  (i.e.,  $p$   
264  $< .05$ ). The same procedure was employed for identifying significant sites across  
265 electrodes in the sensor-space searchlight analysis. For assessing the significance of  
266 effects at the overall classification peaks, repeated-measures ANOVAs and paired t-tests  
267 were performed.

## 268 **3 Results**

### 269 *3.1 Temporal dynamics of pairwise object classification*

270 In a first step, we characterized the overall response dynamics observed in the  
271 pairwise classification analysis, which allowed us to restrict subsequent analyses to time  
272 points where classification performance was particularly pronounced. For this, we  
273 computed an overall measure of pairwise classification by averaging across all unique off-  
274 diagonal elements of the pairwise classification matrices (Figure 2C), resulting in a single  
275 classification time series. This analysis revealed robust above-chance classification  
276 starting from 70 ms after stimulus onset and prominently peaking at 140 ms and 220 ms  
277 (Figure 3A). These two clear peaks in the classification time series were used as time  
278 points of interest for subsequent analyses, as we reasoned that differences between  
279 typically and atypically positioned objects would be most pronounced at time points at  
280 which objects were most discriminable.



281

282 **Figure 3.** Classification Results. A) Overall pairwise classification performance was  
283 computed by averaging all pairwise decoding time series, revealing significant decoding  
284 accuracy starting at 70 ms after stimulus onset and peaking at 140 ms and 220 ms.  
285 Asterisks above data curves indicate above-chance classification ( $p < .05$ , corrected for  
286 multiple comparisons). B) Classification time series for typically and atypically positioned  
287 objects were computed by averaging all pairwise classification time series comparing  
288 typically and atypically positioned objects, respectively. Classification of typically and  
289 atypically positioned pairs showed comparable temporal dynamics, both peaking at the

290 time points identified in the overall decoding. Asterisks indicate above-chance  
291 classification ( $p < .05$ , corrected for multiple comparisons). C) At the first decoding peak  
292 (140 ms), but not the second peak (220 ms), classification was more accurate for typically  
293 than for atypically positioned objects. The asterisk indicates a significant difference ( $p <$   
294  $.05$ ). Error bars reflect standard errors. D) A sensor-space classification searchlight  
295 revealed that the regularity effect (difference between the classification of typically and  
296 atypically positioned objects) at the 140 ms peak is most pronounced in occipital and  
297 temporal electrodes. Circles indicate electrodes exhibiting a significant regularity effect  
298 ( $p < .05$ , corrected for multiple comparisons).

299

### 300 *3.2 Classification of objects when positioned typically and atypically*

301 To test whether neural representations differ for typically and atypically  
302 positioned objects, we compared classification performance for all typically and all  
303 atypically positioned objects. We averaged all pairwise classification time courses that  
304 corresponded to comparisons within regular pairs (e.g., an airplane in the upper visual  
305 field versus a boat in the lower visual field) and comparisons within irregular pairs (e.g.,  
306 an airplane in the lower visual field versus a boat in the upper visual field) (Figure 3B). The  
307 classification time series for typically and atypically positioned objects showed a similar  
308 temporal structure and both replicated the peak structure observed in the overall  
309 decoding, allowing for a meaningful comparison between typically and atypically  
310 positioned objects at the classification peaks. We thus restricted statistical comparisons  
311 to two time points of interest: the peak times observed in the overall decoding (140 ms  
312 and 220 ms). For the early peak at 140ms, we found higher classification accuracy for  
313 typically than for atypically positioned objects,  $t(33) = 3.04$ ,  $p = .005$ , while for the later

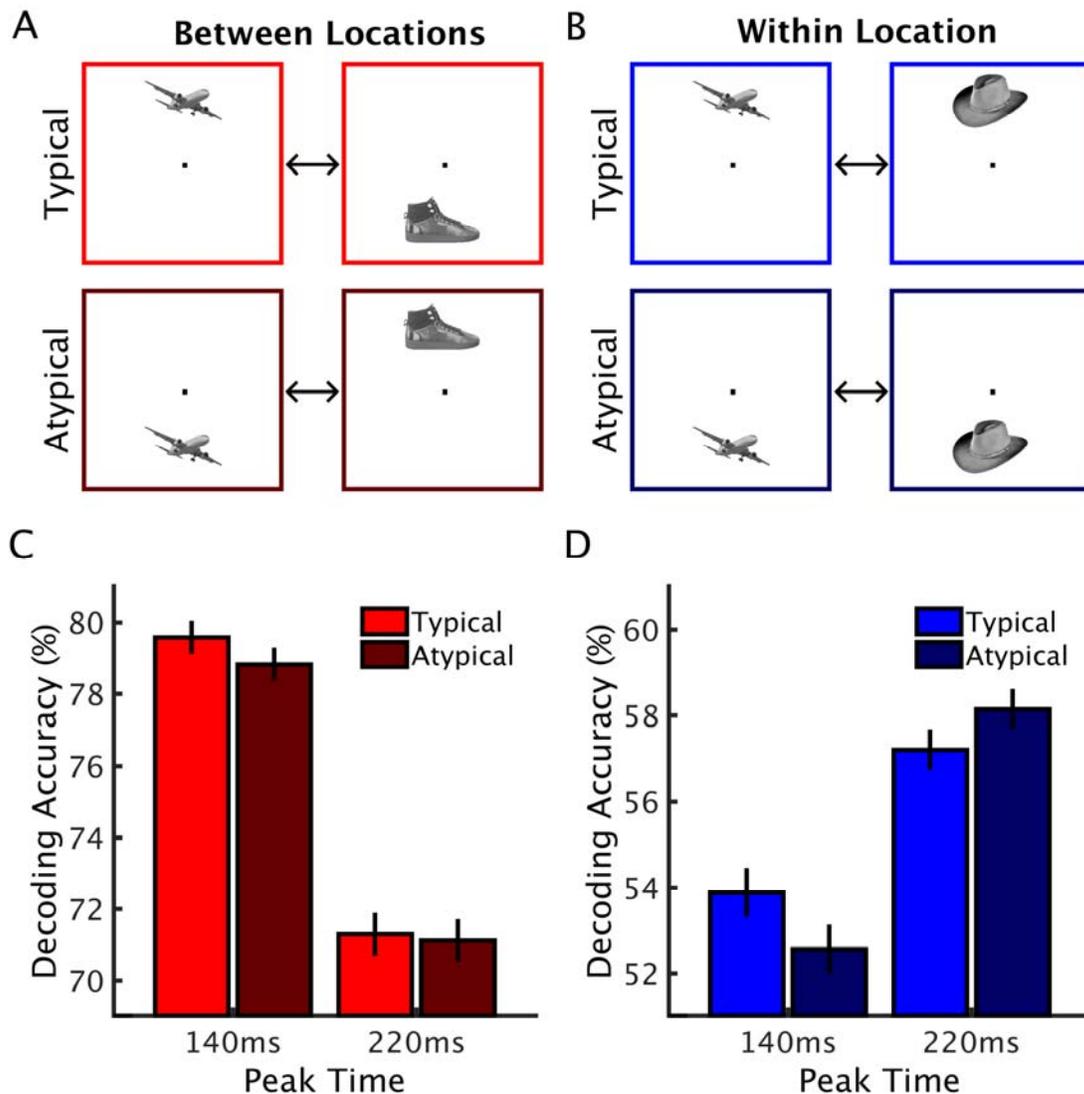
314 peak at 220ms, no such difference emerged,  $t(33) = 0.69$ ,  $p = .495$ , interaction with peak  
315 time,  $F(1,33) = 7.44$ ,  $p = .010$  (Figure 3C). This pattern of results suggests that earlier  
316 stages of object processing (as reflected in the decoding peak at 140 ms) benefit from  
317 typical object locations, while relatively later object representations (at 220 ms after  
318 stimulus onset) are not sensitive to positional regularities.

319 To estimate the spatial extent of the early regularity effect, we performed a  
320 searchlight analysis in sensor space. We repeatedly performed the pairwise classification  
321 analysis for neighborhoods of seven adjacent sensors, using only data from the early  
322 peak at 140 ms. To quantify the regularity benefit, we then computed the difference  
323 between all pairwise comparisons of typically positioned objects and all pairwise  
324 comparisons of atypically positioned objects at every sensor location. This analysis  
325 revealed a significant regularity effect in posterior and lateral electrodes (19 significant  
326 electrode sites) (Figure 3D), suggesting that the enhanced classification for regularly  
327 positioned objects originates from visual areas of the occipital and temporal cortices.

### 328 *3.3 Classification within and between locations*

329 Our classification approach collapsed across pairwise comparisons within the  
330 same location and between different locations, so that classifiers could rely on  
331 information from both an object's identity and its location. A recent study on location  
332 priors in face-part processing (de Haas et al., 2016) has found effects of positional  
333 regularities only when comparing response patterns between locations. To investigate  
334 whether also the regularity effects for objects are differently pronounced when  
335 classifying between the two locations or within the same location, we separately looked  
336 at the regularity effect for all comparisons between locations (e.g., an airplane in the  
337 upper visual field versus a shoe in the lower visual field) (Figure 4A) and all comparisons

338 within location (e.g., an airplane in the upper visual field versus a hat in the upper visual  
339 field) (Figure 4B). We found a main effect of visual field comparison,  $F(1,33) = 365.70$ ,  $p <$   
340  $.001$ , with higher classification accuracies for classifying between locations (where the  
341 classifier can use the stimulus' location) than within location (where the classifier has no  
342 location information available), and an interaction of the within-between comparison and  
343 peak latency,  $F(1,33) = 72.30$ ,  $p < .001$ , with a relatively more pronounced early peak when  
344 classifying between locations. Replicating our previous results, the analysis produced a  
345 significant peak X regularity interaction,  $F(1,33) = 9.83$ ,  $p = .004$ , with a regularity benefit  
346 at the 140 ms peak,  $t(33) = 3.15$ ,  $p = .004$ , but not the 220 ms peak,  $t(33) = 1.05$ ,  $p = .301$ .  
347 Crucially, this pattern of results did not depend on the type of classification (between  
348 locations versus within location),  $F(1,33) = 2.53$ ,  $p = .121$ , suggesting that typical real-world  
349 locations comparably boost early object classification when objects appear in similar or  
350 different locations (Figure 4C,D).



351

352 **Figure 4.** Between-Locations versus Within-Location Classification. We compared peak  
353 decoding for typically and atypically positioned objects separately for comparisons  
354 between different locations (e.g., an airplane in the upper visual field versus a shoe in the  
355 lower visual field) (A) and within the same location (e.g., an airplane in the upper visual  
356 field versus a hat in the upper visual field) (B). For both comparison types, we found a  
357 similar pattern (C, D) with a benefit for typically positioned objects at the 140 ms peak.  
358 Importantly, the patterns of results for between-locations and within-location  
359 classification were statistically indistinguishable (see Results). Error bars reflect standard  
360 errors.

## 361 **4 Discussion**

### 362 *4.1 Summary*

363 Here, we demonstrate that positional regularities contained in real-world scenes  
364 impact brain responses to individual objects. Using multivariate classification of EEG data,  
365 we show that object coding across the visual field is affected by the typical real-world  
366 location of the object. When objects are presented in frequently experienced locations,  
367 EEG response patterns at 140 ms after stimulus onset are better discriminable than when  
368 the same objects are presented in atypical locations. This advantage for typically  
369 positioned objects was equally pronounced for classification between locations and  
370 within the same location. Using a sensor-space searchlight analysis, we show that the  
371 effect was primarily localized in posterior and lateral sensors, suggesting that typically  
372 positioned objects gain an advantage during early perceptual processing.

### 373 *4.2 Early stages of object coding are sensitive to typical locations*

374 Our results support the hypothesis that extensive experience with natural scene  
375 structure can enhance object coding when objects appear in typical locations of the visual  
376 field. This finding demonstrates that visual processing channels are preferentially tuned  
377 to specific objects appearing in specific locations (Kaiser & Haselhuhn, 2017; Kravitz et al.,  
378 2008). Crucially, our EEG classification approach allowed us to pinpoint the latency of this  
379 regularity benefit: We demonstrate that object processing at 140 ms after stimulus onset  
380 is affected by positional regularity. The timing of the effect suggests that objects  
381 appearing in typical visual field locations gain an advantage during early, perceptual  
382 processing, rather than through top-down interactions within visual cortex or feedback  
383 from frontal areas; previous M/EEG studies have suggested that such feedback processes  
384 impact visual responses only at later stages, starting shortly before 200 ms (Bar et al.,

385 2006; Fahrenfort, van Leeuwen, Olivers, & Hogendoorn, 2017). As opposed to the  
386 difference in early object processing, later representations (at 220 ms after stimulus  
387 onset) do not depend on the location of the object. This result concurs with the  
388 increasing location tolerance over the time course of object classification, peaking at  
389 around 180 ms (Isik et al., 2014), mirroring the increase in receptive field size along the  
390 visual stream (Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013).

#### 391 *4.3 Visual versus categorical sources of the regularity benefit*

392 What is the content of the location-specific object representations emerging at  
393 140 ms? Peaks in the M/EEG decoding in this time range have been previously associated  
394 with visual category processing in object-selective cortex (Cichy, Pantazis, & Oliva, 2014,  
395 2016; Carlson et al., 2013). Our searchlight analysis reveals the strongest regularity effect  
396 over lateral occipital and temporal electrode sites, suggesting that the effect originates  
397 from object-selective visual cortex. Whether processing differences in these object-  
398 selective regions reflect genuine category processing differences or whether they reflect  
399 differential coding of category-associated visual features is a debated question (Bracci,  
400 Ritchie, & op de Beeck, 2017; Peelen & Downing, 2017). While some data suggest that  
401 visual properties explain most of the variance in object-selective responses (e.g., Baldassi  
402 et al., 2013), a recent MEG decoding study has demonstrated category-selective, rather  
403 than visually (shape-) driven, responses from as early as 130 ms after stimulus onset  
404 (Kaiser, Azzalini, & Peelen, 2016). To determine if the regularity benefit observed here can  
405 be linked to differences in the processing of particular visual features or true categorical  
406 processing differences, future studies need to employ stimuli that vary more extensively  
407 in their visual characteristics.

#### 408 *4.4 Positional structures in multiple and individual objects*

409           Natural environments contain positional regularities on different levels, both on  
410 the levels of multiple (e.g., a lamps typically hang above tables) and individual objects  
411 (e.g., a lamp is typically in the upper visual field). Previous research has primarily focused  
412 on the latter: Recent behavioral studies have demonstrated that regularity structures in  
413 multi-object arrangement facilitate behavior in capacity-limited visual tasks (Gronau &  
414 Shachar, 2014; Kaiser et al., 2014; Kaiser, Stein, & Peelen, 2015; Stein, Kaiser, & Peelen,  
415 2015), and neuroimaging studies demonstrated that they enable the brain to integrate  
416 information across objects that appear in frequently experienced arrangements (Baeck,  
417 Wagemans, & Op de Beeck, 2013; Kaiser & Peelen, 2017; Kaiser et al., 2014).

418           Here, we provide the first evidence that typical regularity structures also impact  
419 the neural representation of individual objects. Our finding thus raises the question  
420 whether the previously reported regularity effects in multi-object perception can be  
421 reduced to the effects of typical individual object location. On a behavioral level, some  
422 previous studies oppose this notion by demonstrating that the benefits of multi-object  
423 regularities cannot be explained by the relative location of the constituent objects (Kaiser  
424 et al., 2014, 2015; Stein et al., 2015). Although these results suggest that positional  
425 regularities in multi-object and single-object processing offer complementary benefits,  
426 further research is needed. Future studies will need to systematically manipulate  
427 positional structures on different levels (from individual objects to multi-object  
428 arrangements) to explore how regularities on multiple levels interact on a neural level.

#### 429 *4.5 Positional structures beyond person perception*

430           Positional regularities have been studied in humans and non-human primates  
431 largely in the context of face and body perception, where parts are arranged in highly  
432 predictable configurations (e.g., the features of a human face). fMRI studies in humans

433 have demonstrated that individual face and body parts are processed more efficiently  
434 when they appear in typical visual field locations (Chan et al., 2010; de Haas et al., 2016).  
435 Single-cell recordings in monkeys demonstrated that location biases can impact cortical  
436 responses to face parts as early as 100 ms after stimulus onset (Issa & DiCarlo, 2012),  
437 suggesting a benefit at early stages of perceptual processing.

438         Our results complement these findings by showing that such location-specific  
439 object processing is not restricted to the face/body domain: The inherent structure of  
440 natural scenes can similarly impact early processing (140 ms after stimulus onset) of  
441 object information across the visual field. Our findings thus highlight that location-specific  
442 tuning in object processing may form a general principle that shapes visual processing  
443 mechanisms for spatially predictable information. Future research could test whether  
444 regularity structures also affect other domains where the visual input consists of multiple  
445 parts that are constrained by spatial regularities. For example, through extensive  
446 experience with reading written text, the neural mechanisms for perceiving letters could  
447 get tuned to their typical spatial locations within words (Kaiser & Haselhuhn, 2017;  
448 Vinckier et al., 2007).

#### 449 *4.6 Location-specific object coding and efficient scene perception*

450         A major challenge for the visual system when processing natural scenes is the  
451 large number of individual objects they contain. The concurrent representation of  
452 multiple objects is limited by an overlap in processing resources, as indexed by reduced  
453 neural responses when multiple objects need to be processed (Cohen, Konkle, Rhee,  
454 Nakayama, & Alvarez, 2014; Desimone & Duncan, 1995; Franconeri, Alvarez, & Cavanagh,  
455 2013). The preferential coding for typically positioned objects revealed here may  
456 contribute to efficient scene processing by reducing this inter-object competition. By

457 coding the objects of a scene via separable location-specific processing channels that are  
458 optimally tuned to objects typically appearing in these locations, the overlap in neural  
459 processing resources can be reduced. Linking this increase in neural efficiency to a  
460 facilitation of perceptual performance could help to understand the efficient processing  
461 of highly complex natural environments.

462 **Acknowledgements**

463 The research was supported by a DFG Emmy-Noether Grant awarded to R.M.C. (CI241-1/1).

464 **References**

- 465 Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2013). The distributed representation of  
466 random and meaningful object pairs in human occipitotemporal cortex: the  
467 weighted average as a general rule. *Neuroimage*, *70*, 37-47.
- 468 Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R., & Zoccolan, D. (2013).  
469 Shape similarity, better than semantic membership, accounts for the structure of  
470 visual object representations in a population of monkey inferotemporal neurons.  
471 *PLoS Computational Biology*, *9*, e1003167.
- 472 Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617-629.
- 473 Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen,  
474 M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down  
475 facilitation of visual recognition. *Proceedings of the National Academy of Sciences*,  
476 *U.S.A.*, *103*, 449-454.
- 477 Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting  
478 and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143-  
479 177.
- 480 Bracci, S., Ritchie, J. B., & op de Beeck, H. P. (2017). On the partnership between neural  
481 representations of object categories and visual features in the ventral visual  
482 pathway. *Neuropsychologia*, doi:10.1016/j.neuropsychologia.2017.06.010.
- 483 Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.
- 484 Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal  
485 resolution decoding of object position and category. *Journal of Vision*, *11*, 9.
- 486 Carlson, T. A., Tovar D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics  
487 of object vision: the first 1000 ms. *Journal of Vision*, *13*, 1-19.

- 488 Chan, A. W., Kravitz, D. J., Truong, S., Arizpe, J., & Baker, C.I. (2010). Cortical  
489 representations of bodies and faces are strongest in commonly experienced  
490 configurations. *Nature Neuroscience*, 13, 417-418.
- 491 Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4,  
492 170-178.
- 493 Cichy, R. M., Chen, Y., & Haynes, J. D. (2011). Encoding the identity and location of objects  
494 in human LOC. *Neuroimage*, 54, 2297-2307.
- 495 Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space  
496 and time. *Nature Neuroscience*, 17, 455-462.
- 497 Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-based fusion of MEG and fMRI  
498 reveals spatio-temporal dynamics in human cortex during visual object  
499 recognition. *Cerebral Cortex*, 26, 3563-3579.
- 500 Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing of  
501 multiple visual objects is limited by overlap in neural channels. *Proceedings of the*  
502 *National Academy of Sciences, U.S.A.*, 111, 8955-8960.
- 503 Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object  
504 recognition in the human brain: From visual features to categorical decisions.  
505 *Neuropsychologia*, doi:10.1016/j.neuropsychologia.2017.02.013.
- 506 de Haas, B., Schwarzkopf, D. S., Alvarez, I., Lawson, R. P., Henriksson, L., Kriegeskorte, N.,  
507 & Rees, G. (2016). Perception and processing of faces in the human brain is tuned  
508 to typical facial feature locations. *Journal of Neuroscience*, 36, 9289-9302.
- 509 Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention.  
510 *Annual Review of Neuroscience*, 18, 193-222.

- 511 Fahrenfort, J. J., van Leeuwen, J., Olivers, C. N. L., & Hogendoorn, H. (2017). Perceptual  
512 integration without conscious access. *Proceedings of the National Academy of*  
513 *Sciences, U.S.A.*, 114, 3744-3749.
- 514 Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources:  
515 competitive content maps for attention and memory. *Trends in Cognitive Sciences*,  
516 17, 134-141.
- 517 Golomb, J. D., & Kanwisher, N. (2012). Higher level visual cortex represents retinotopic,  
518 not spatiotopic, object location. *Cerebral Cortex*, 22, 2794-2810.
- 519 Gronau, N., & Shachar, M. (2014). Contextual integration of visual objects necessitates  
520 attention. *Attention, Perception & Psychophysics*, 76, 695-714.
- 521 Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an  
522 organizing principle for human high-order object areas. *Neuron*, 34, 479-490.
- 523 Hemond, C. C., Kanwisher, N., & Op de Beeck, H. P. (2007). A preference for contralateral  
524 stimuli in human object- and face-selective cortex. *PLoS One*, 2, e574.
- 525 Henriksson, L., Mur, M., & Kriegeskorte, N. (2015). Faciotopy – A face-feature like  
526 topology in the human occipital face area. *Cortex*, 72, 156-167.
- 527 Hong, H, Yamins, D. L. K., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for  
528 category-orthogonal object properties increases along the ventral visual stream.  
529 *Nature Neuroscience*, 19, 613-622.
- 530 Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object  
531 recognition in the human visual system. *Journal of Neurophysiology*, 111, 91-102.
- 532 Issa, E. B., & DiCarlo, J. J. (2012). Precedence of the eye region in neural processing of  
533 faces. *Journal of Neuroscience*, 32, 16666-16682.

- 534 Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category  
535 responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, *115*,  
536 2246-2250.
- 537 Kaiser, D., & Haselhuhn, T. (2017). Facing a regular world: How spatial object structure  
538 shapes visual processing. *Journal of Neuroscience*, *37*, 1965-1967.
- 539 Kaiser, D., & Peelen, M. V. (2017). Transformation from independent to integrative coding  
540 of multi-object arrangements in human visual cortex. *BioRxiv*, doi:10.1101/117432.
- 541 Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world  
542 regularities facilitates perception by reducing competitive interactions in visual  
543 cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, 11217-11222.
- 544 Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world spatial regularities affect visual  
545 working memory for objects. *Psychonomic Bulletin & Review*, *22*, 1784-1790.
- 546 Kravitz, D. J., Kriegeskorte, N., & Baker, C. I. (2010). High-level visual object  
547 representations are constrained by position. *Cerebral Cortex*, *20*, 2916-2925.
- 548 Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The  
549 ventral visual pathway: an expanded neural framework for the processing of  
550 object quality. *Trends in Cognitive Sciences*, *17*, 26-49.
- 551 Kravitz, D. J., Vinson, L. D., & Baker, C. I. (2008). How position dependent is visual object  
552 recognition? *Trends in Cognitive Sciences*, *12*, 114-122.
- 553 Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in*  
554 *Cognitive Sciences*, *11*, 520-527.
- 555 Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). Fieldtrip: open source  
556 software for advances analysis of MEG, EEG, and invasive electrophysiological  
557 data. *Computational Intelligence and Neuroscience*, *2011*, 156869.

- 558 Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal  
559 multivariate pattern analysis of neuroimaging data in Matlab / GNU Octave.  
560 *Frontiers in Neuroinformatics*, 10, 20.
- 561 Peelen, M. V., & Downing, P. E. (2017) Category selectivity in human visual cortex: Beyond  
562 visual object recognition. *Neuropsychologia*,  
563 doi:10.1016/j.neuropsychologia.2017.03.033.
- 564 Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective  
565 Neuroscience*, 2, 67-70.
- 566 Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database  
567 and web-based tool for image annotation. *International Journal of Computer Vision*,  
568 77, 157-173.
- 569 Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of  
570 category and location information across object-selective regions in human visual  
571 cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, 105, 4447-4452.
- 572 Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing  
573 problems of smoothing, threshold dependence and localization in cluster  
574 interference. *Neuroimage*, 66, 215-222.
- 575 Stein, T., Kaiser, D., & Peelen, M. V. (2015). Interobject grouping facilitates visual  
576 awareness. *Journal of Vision*, 15, 10.
- 577 Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007).  
578 Hierarchical coding of letter strings in the ventral stream: dissecting the inner  
579 organization of the visual word-form system. *Neuron*, 55, 143-156.

- 580 Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010).  
581 Controlling low-level image properties: The SHINE toolbox. *Behavior Research*  
582 *Methods*, 42, 671-684.
- 583 Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes  
584 involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 15, 77-  
585 84.