

1

2

3

4

5 **Decoding images in the mind's eye: The temporal dynamics of visual imagery**

6

7 Sophia M. Shatek <sup>1,\*</sup>, Tijl Grootswagers <sup>1,2</sup>, Amanda K. Robinson<sup>1,2,3</sup>, Thomas A. Carlson<sup>1,2</sup>

8

9 <sup>1</sup> School of Psychology, University of Sydney , Australia

10 <sup>2</sup> ARC Centre of Excellence in Cognition & its Disorders, Macquarie University, Sydney,  
11 Australia

12 <sup>3</sup> Perception in Action Research Centre & Department of Cognitive Science, Macquarie  
13 University, Sydney, Australia

14

15 \* Corresponding author: Sophia Shatek

16 Email: [sophia.shatek@sydney.edu.au](mailto:sophia.shatek@sydney.edu.au)

17

18

19

20

21

22

23

24

## 25 **Abstract:**

26 Mental imagery is the ability to generate images in the mind in the absence of sensory  
27 input. Both perceptual visual processing and internally generated imagery engage large,  
28 overlapping networks of brain regions. However, it is unclear whether they are  
29 characterized by similar temporal dynamics. Recent magnetoencephalography work has  
30 shown that object category information was decodable from brain activity during mental  
31 imagery, but the timing was delayed relative to perception. The current study builds on  
32 these findings, using electroencephalography to investigate the dynamics of mental  
33 imagery. Sixteen participants viewed two images of the Sydney Harbour Bridge and two  
34 images of Santa Claus. On each trial, they viewed a sequence of the four images and were  
35 asked to imagine one of them, which was cued retroactively by its temporal location in the  
36 sequence. Time-resolved multivariate pattern analysis was used to decode the viewed and  
37 imagined stimuli. Our results indicate that the dynamics of imagery processes are more  
38 variable across, and within, participants compared to perception of physical stimuli.  
39 Although category and exemplar information was decodable for viewed stimuli, there were  
40 no informative patterns of activity during mental imagery. The current findings suggest  
41 stimulus complexity, task design and individual differences may influence the ability to  
42 successfully decode imagined images. We discuss the implications of these results for our  
43 understanding of the neural processes underlying mental imagery.

44 **Keywords:** mental imagery; electroencephalography; MVPA; decoding

45

46

47

48

## 49 Introduction

50 Does the Mona Lisa face left or right? A common method of solving this problem is to  
51 form an image of the Da Vinci painting in your 'mind's eye'. Our ability to imagine scenes  
52 and objects can help us solve everyday problems and accomplish day-to-day tasks, such  
53 as retracing our steps to find a lost item or navigating from a memorised map. These  
54 mentally-generated images are formed in the absence of visual information, and are instead  
55 based on short- or long-term memories (Ganis et al., 2003; Kosslyn et al., 2001). Images  
56 generated from memory seem anecdotally weaker, or less vivid, than those evoked by  
57 sensory input, yet also appear to rely on the visual system (Dijkstra et al., 2018). In line with  
58 this, current theories of mental imagery involve common mechanisms for human vision and  
59 mental imagery.

60 Recent work has revealed overlapping neural substrates for visual perception and  
61 imagery. Positron emission tomography (PET) and functional magnetic resonance imaging  
62 (fMRI) have revealed similar patterns of brain activity during perception and imagery,  
63 suggesting computational overlap in the neural systems responsible for each process (Ganis  
64 et al., 2004; Kosslyn et al., 1999; Lee et al., 2012; Slotnick et al., 2005). This overlap is  
65 particularly clear for areas associated with higher-order abstract visual processing, such as  
66 visual association cortex (Albers et al., 2013; Goldenberg et al., 1989; Knauff et al., 2000)  
67 and category-selective temporal cortices (Mechelli et al., 2004; Reeder et al., 2015).  
68 Overlapping activation is also present in low-level visual areas, despite the absence of visual  
69 input during imagery; imagery and visual perception both activate the lateral geniculate  
70 nucleus of the thalamus (LGN) (Chen et al., 1998) and primary visual cortex (V1) (Albers et  
71 al., 2013; Harrison and Tong, 2009; Pearson et al., 2008). Together, this supports the notion  
72 that imagery utilises many of the same mechanisms as visual perception.

73 Despite overlapping neural activation for vision and imagery, the neural processes are  
74 not identical. For example, there is more overlap in higher, anterior regions (i.e., frontal and  
75 parietal; Ganis et al., 2004), compared to lower, posterior visual regions (Harrison and Tong,

76 2009; Lee et al., 2012). There are also task-related differences in imagery such that different  
77 imagery tasks show varying degrees of overlap with vision (Ganis et al., 2004; Ishai et al.,  
78 2000; Kosslyn and Thompson, 2003). Patients with brain damage also provide evidence for  
79 dissociation between imagery and vision. Some patients with occipital or parietal lesions can  
80 successfully complete tasks relying on mental imagery, despite significant visual deficits,  
81 while others have fully functioning vision but impaired imagery (Bartolomeo et al., 2013;  
82 Bridge et al., 2012; Moro et al., 2008; Zago et al., 2010). Therefore, there is some  
83 dissociation between vision and imagery despite similar neural processing.

84 To date, research has focused on understanding the brain networks recruited by a  
85 variety of imagery tasks (Fulford et al., 2018; Mechelli et al., 2004), yet we have very little  
86 understanding of the temporal dynamics of mental imagery. Although fMRI studies have  
87 found correlations between imagery and perception in the later stages of visual processing  
88 (Stokes et al., 2011), as well as similar activation patterns between imagery and working  
89 memory (Albers et al., 2013), this evidence is limited by the temporal resolution of fMRI.  
90 Recent work using MEG has revealed that while similar activation patterns are present in  
91 imagery and vision, they occur at a later time and are more diffuse, pointing towards a  
92 temporal dissociation between the two seemingly similar processes (Dijkstra et al., 2018).

93 Multi-Variate Pattern Analysis (MVPA) applied to neuroimaging data can elucidate the  
94 information represented in different brain regions (fMRI), and at particular points in time  
95 (M/EEG). MVPA offers an advantage in analysing data from mental imagery, as analyses  
96 are conducted at an individual-subject level and mental imagery ability is understood to vary  
97 significantly between people (e.g., Cui et al., 2007). MVPA is also more sensitive to variation  
98 across fine-grained patterns, and provides a powerful framework for the detection of content-  
99 specific information (Grootswagers et al., 2017; Haynes, 2015). This is particularly  
100 advantageous for imagery signals that are likely to be weaker than visual input (Naselaris et  
101 al., 2015). One recent study found that the category of imagined images (faces and houses)  
102 was decodable from MEG recordings, albeit later than viewed images (Dijkstra et al., 2018).

103 However, decoding of individual exemplars was poor, indicating a dissociation between low-  
104 and high-level imagery processes.

105 Here, we examined how the neural representation of mental images develops and  
106 changes over time. Participants imagined one of four previously learned pictures: two faces  
107 and two places. Each image was visually dissimilar to the other within the category, while  
108 maintaining clear category divisions. Neural responses were measured using EEG while  
109 participants viewed the experimental images, imagined the images, and viewed fast streams  
110 of semantically related images (i.e., other faces and places). We expected that category  
111 information would be decodable from the EEG data during mental imagery (Dijkstra et al.,  
112 2018), that it would be broadly generalisable across the imagery period, and delayed relative  
113 to vision. We also predicted that exemplars within each category would be distinguishable  
114 (i.e., successful within-category decoding). We found that the dynamics of imagery  
115 processes are more variable across, and within, participants compared to perception of  
116 physical stimuli. Although category and exemplar information was decodable for viewed  
117 stimuli, there were no informative patterns of activity during mental imagery.

118

## 119 **Materials and Methods**

### 120 **Experimental structure**

121 At the start of the session, participants completed the Vividness of Visual Imagery  
122 Questionnaire (VVIQ) (Marks, 1973). They were then informed of the task instructions and  
123 completed 24 imagery task training trials. The experiment itself consisted of four blocks that  
124 were completed while EEG was measured. In each block, participants passively viewed five  
125 rapid streams of images (Pattern Estimator), followed by a series of imagery trials. Each  
126 imagery trial consisted of a four-image sequence (Seen images), after which participants were  
127 cued to imagine one of those stimuli (Imagery).

128

## 129 **Participants**

130 We recruited 16 right-handed subjects (11 male), of mean age 23 (SD= 5.58, range 18-  
131 39), with normal or corrected-to-normal vision and no history of psychiatric or neurological  
132 disorders. The experiment was approved by the Human Ethics Committee of the University of  
133 Sydney. Written, informed consent was obtained from all participants.

134

## 135 **Behavioural data**

136 To measure individual variation in vividness, we administered a modified VVIQ (Marks,  
137 1973) prior to EEG set-up. The VVIQ measures subjective perception of the strength of an  
138 individual's mental imagery. Participants were asked to imagine 16 scenarios, and rated each  
139 for vividness on a five-point Likert-like scale. A reversed scoring system was used to decrease  
140 confusion. Participants rated each item from 1 (*"No image at all, you only 'know' that you are*  
141 *thinking of an object"*) to 5 (*"Perfectly clear and as vivid as normal vision"*). All questions were  
142 completed twice, once with open eyes and once with closed eyes. A final summed score  
143 between 32 and 160 was calculated for each subject; higher scores indicate greater vividness.

144

## 145 **Apparatus and Stimuli**

146 Four stimuli were used in this experiment: two images of Santa and two images of the  
147 Sydney Harbour Bridge. The inclusion of two exemplars per category allowed us to  
148 disentangle whether participants are thinking of the concept (i.e., Santa, Sydney Harbour  
149 Bridge) or generating a specific image. These stimuli also fit into distinct face/place categories,  
150 which have been shown to evoke robustly distinct patterns of neural activity (Haxby et al.,  
151 2001; Kanwisher et al., 1997).

152 All stimuli were displayed on a 1920 x 1080 pixel Asus monitor on a grey background.  
153 Participants viewed stimuli at approximately 57cm, such that all stimuli subtended  
154 approximately 4.1 degrees of visual angle (including a 0.15 degree black border). Responses  
155 were made using a mouse with the right hand. A grey fixation cross was superimposed on all

156 stimuli, with horizontal and vertical arms subtending approximately 0.6 degrees of visual angle.  
157 Experimental presentations were coded in MATLAB using extensions from the PsychoPhysics  
158 Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

159

## 160 **Imagery sequence**

161 Each imagery sequence began with a fixation cross in the centre of the screen for 1000  
162 milliseconds. The four stimuli were displayed sequentially in the centre of the screen, within a  
163 black border. Each was displayed for 1500 milliseconds each, in a pseudo-random order.  
164 Targets were counterbalanced such that each block contained all 24 possible sequences of  
165 the four stimuli. For each sequence, a different target was selected in each block. Target  
166 allocation in each block was also randomised. This counterbalancing meant each image  
167 appeared in each temporal position as a target equally often.

168 The fourth stimulus was followed by a 1000ms fixation cross, then a numerical cue  
169 appeared (1-4). This cue referred to the target's position in the stream; for example, '3'  
170 indicated the target was the third image in the stream. Participants were instructed to click the  
171 mouse once they had identified the target and were mentally "projecting an image into the  
172 square". Upon clicking, the number was replaced with a dark grey fixation cross and the frame  
173 was filled light grey. This 'imagery' screen was displayed for 3000ms before automatically  
174 advancing to a response screen. On the response screen, participants were shown the four  
175 stimuli and horizontal mirror images of these stimuli. They used a mouse to select which of  
176 these images they were imagining. Mirror images were used as distractors because they are  
177 semantically identical but visually different, to determine if participants were using a semantic  
178 strategy rather than an imagery-based strategy. Horizontal positioning changed across blocks  
179 (stimulus identity), and vertical positioning was randomised every trial (mirror images/stimulus)  
180 such that for some trials the mirror image was in the top row, and some in the bottom row.  
181 This randomisation aimed to reduce predictability in responses.

182

## 183 **Training**

184 Participants completed a block of 24 practice trials of the imagery sequence before EEG  
185 recording. We expected these training trials to give participants the opportunity to learn task  
186 structure and observe more details about the images to facilitate vivid imagery. Training trials  
187 were similar to experimental trials. The first 12 trials contained typed instructions on how to  
188 identify the target, and went straight to the response screen after the cue, with no imagery  
189 component. On incorrect responses, participants were shown the correct image. The second  
190 12 trials mimicked experimental trials, with the addition of typed instructions and feedback.  
191 Participants were given the option to repeat the training, and two did so.

192

## 193 **Pattern Estimator**

194 We also included a pattern estimator at the beginning of each to investigate the degree  
195 of generalisation across semantic category. These images were semantically similar to the  
196 critical experimental stimuli. Participants passively viewed a rapid stream containing the four  
197 stimuli from the imagery sequence, as well as horizontally flipped, inverted and blurred  
198 versions of these images. It also included other images of the Sydney Harbour Bridge and  
199 Santa, other bridges and other people. Each block began with five short streams of 56 images,  
200 displayed for 200ms each. Every stream contained all 56 images in a random order, and lasted  
201 for 11.2 seconds. Participants could pause between streams and elected to advance when  
202 they were ready.

203

## 204 **Data recording and processing**

### 205 **EEG recording**

206 EEG data were continuously recorded at 1000Hz using a 64-channel Brain Products  
207 (GmbH, Herrsching, Germany) ActiCAP system with active electrodes. Electrode locations  
208 corresponded to the modified 10-10 international system for electrode placement (Oostenveld



209 and Praamstra, 2001), with the online reference at Cz. Electrolyte gel kept impedances below  
210 10k $\Omega$ .

211

## 212 **Pre-processing EEG**

213 EEG pre-processing was completed offline using EEGLAB (Delorme and Makeig, 2004)  
214 and ERPLAB (Lopez-Calderon and Luck, 2014). The data were minimally pre-processed.  
215 Data were down-sampled to 250Hz to reduce computational load, then filtered using a 0.1Hz  
216 high-pass filter, and a 100Hz low-pass filter. Line noise at 50Hz was removed using the  
217 CleanLine function in EEGLAB. Four types of epochs were created: Pattern Estimator, Vision,  
218 Cue-Locked Imagined and Response-Locked Imagined. Each epoch included 300ms before  
219 to 1500ms after stimulus onset. Pattern Estimator epochs were from the fast stream at the  
220 beginning of each block, and Vision epochs were from the four images displayed in each  
221 experimental trial. Cue-locked Imagined epochs were centred around presentation of the  
222 numerical cue designating the target. Response-Locked Imagined epochs were centred  
223 around participants' mouse click to begin imagery. Although the period between cue and  
224 response was variable across trials (Supplementary Fig S2), we expected the period  
225 immediately following the cue to provide insight into the initial stages of imagery generation.

226

## 227 **Decoding analysis**

228 All EEG analyses were performed using time-resolved decoding methods, custom-written  
229 using CoSMoMVPA functions in MATLAB (Oosterhof et al., 2016). For all decoding analyses,  
230 a regularised linear discriminant classifier (as implemented in CoSMoMVPA) was trained to  
231 differentiate brain patterns evoked by each image or category of images.

232 For category decoding, a classifier was trained to distinguish images of Santa from  
233 images of the Sydney Harbour Bridge for recordings from the same type (i.e., a classifier  
234 trained on data from the Pattern Estimator was tested on another independent portion of the  
235 Pattern Estimator data). To determine if exemplars were also uniquely represented, a classifier  
236 was trained to distinguish between the two exemplars within each category (e.g., decode the

237 two Santa images). Classifiers were trained and tested for each time point using a 12ms sliding  
238 time window (three time points).

239 To analyse data from the Pattern Estimator and Vision epochs, each presentation  
240 sequence was treated as independent. We used a leave-one-trial-out cross-validation  
241 approach, where Vision trials were composed of the four stimuli in each imagery sequence  
242 and Pattern Estimator trials were composed of a single sequence containing all 56  
243 semantically relevant images. Imagined stimuli were analysed using a leave-two-out cross-  
244 validation approach, which took each imagery epoch as independent and left one exemplar of  
245 each category (one Santa and one Sydney Harbour Bridge) in the test set. Cross-decoding  
246 analyses were conducted using split-half cross-validation, where a classifier was trained on  
247 one trial type and tested on another trial type (e.g., train on all Vision trials and test on all Cue-  
248 Locked Imagined trials). To investigate the possibility of similar processes occurring in vision  
249 and imagery at different times, we used temporal generalisation methods (King and Dehaene,  
250 2014), in which the trained classifier for a single time point is applied to every time point in a  
251 second set of data.

252 To compute statistical probability for all within-type, cross-decoding and time  
253 generalisation analyses, we used the Monte Carlo Cluster Statistics function in the  
254 CoSMoMMPA toolbox (Maris and Oostenveld, 2007; Smith and Nichols, 2009; Stelzer et al.,  
255 2013). These statistics yield a corrected p-value that represents the chance that the decoding  
256 accuracy could have come from a null distribution formed from 10,000 iterations (North et al.,  
257 2002). These p-values were thresholded at  $p_{\text{corrected}} < .05$  for significance.

258

## 259 **Results**

260 In this experiment, participants viewed rapid streams of images (Pattern Estimator), and  
261 series of imagery trials. In imagery trials, participants were presented with a sequence of four  
262 images (Vision) and then were cued to imagine one of the images (Imagery). We trained and  
263 tested multivariate classifiers to decode exemplar and category of the object in all three

264 conditions, as well as tested the generalisation performance of classifiers between vision and  
265 imagery trials.

## 266 **Behavioural results**

### 267 **Vividness of Visual Imagery Questionnaire**

268 The VVIQ was scored out of 160, a sum of responses to each of the 16 questions on a  
269 five-point scale. The VVIQ was given to participants both with eyes open and closed (Marks,  
270 1973). The average overall score was 113 ( $SD = 15.93$ , range 82-150), similar to previously  
271 reported means (Amedi et al., 2005; Crawford, 1982; Fulford et al., 2018). Responses with  
272 eyes open ( $M = 56.44$ ,  $SD = 8.54$ ) were very similar to eyes closed ( $M = 57.69$ ,  $SD = 10.28$ ).  
273 The distribution of overall scores is shown in Supplementary Figure S1.

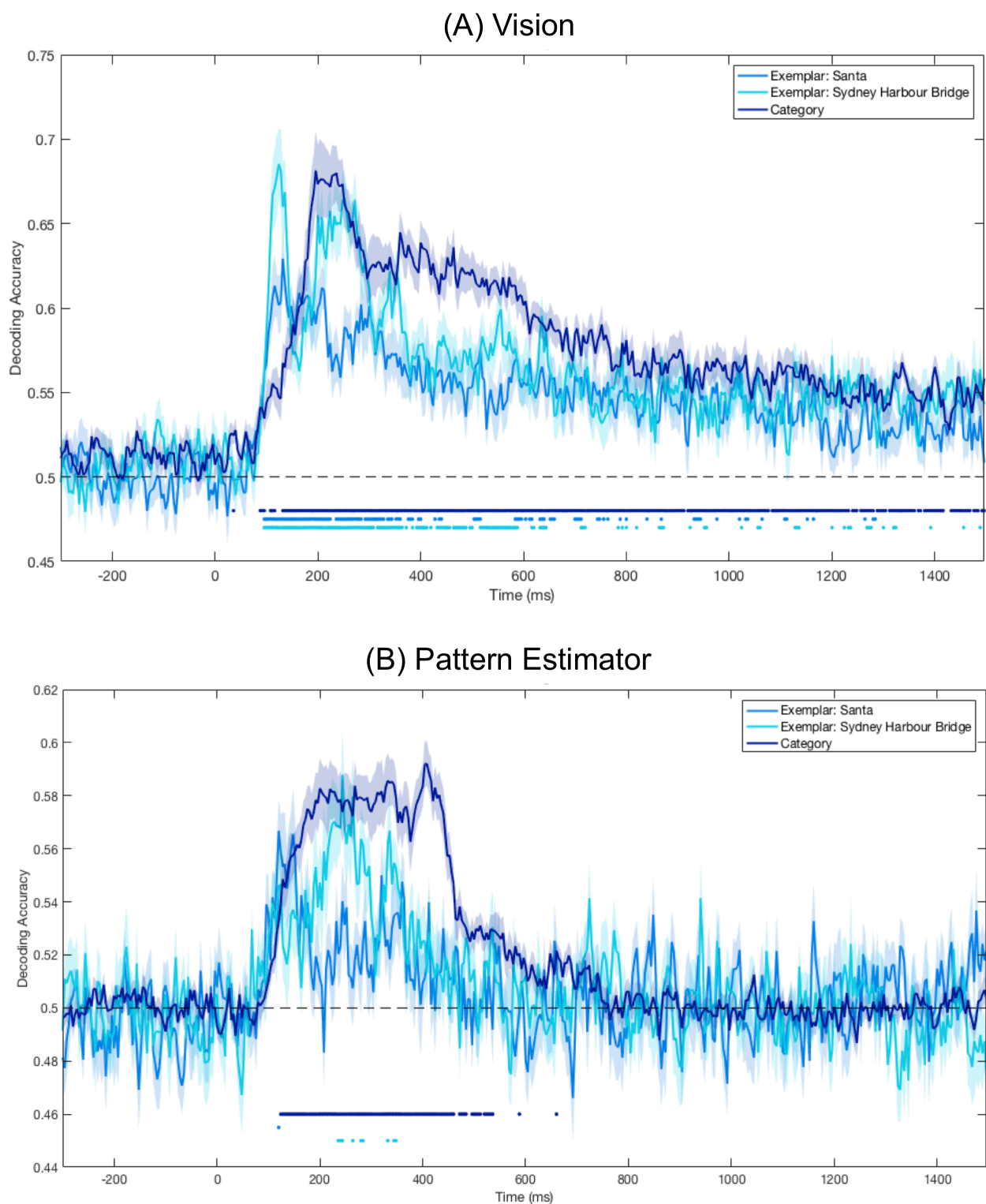
### 274 **Target identification**

275 To verify if participants were able to identify the target for imagery trials correctly, we  
276 examined their behavioural responses after each imagery sequence. Participants were able  
277 to accurately identify the target, with an average overall accuracy of 92% ( $SD = 4.40$ ). Of  
278 the trials which were errors, most participants chose one of the four original images (67% of  
279 errors). Approximately a third of incorrect responses were to the flipped version of the target.  
280 This suggests participants successfully learned the basic characteristics of the target images  
281 and were not simply relying on a mnemonic strategy to complete the task. The mean  
282 response time from cue to imagery was 3.21 seconds ( $SD = 1.86$ ) and the most frequent  
283 response time was between 1.5 and 2 seconds (Supplementary Fig S2).

## 284 **EEG results**

285 **Significant decoding of image category and exemplars for seen**  
286 **images on imagery trials**

287 To test whether category information was represented in visually displayed images, we  
288 trained and tested a classifier on the images seen during experimental trials (Vision).  
289 Category decoding was continuously above chance ( $p < .05$ ) after 88ms (Fig 1), indicating  
290 patterns of brain activity for Santas and Sydney Harbour Bridges were distinguishable from  
291 this point. This above-chance decoding was sustained for the entire time the image was  
292 displayed. Continuous above-chance decoding began for both Santas and Sydney Harbour  
293 Bridges at 96ms. Peak accuracy occurred at 132ms for Santas, 124ms for Sydney Harbour  
294 Bridges and at 196ms for category decoding.



**Figure 1.** Mean decoding accuracy for Vision (A) and Pattern Estimator (B) images. Dots below plots indicate time points at which decoding was significantly above chance ( $p < .05$ ). Shaded areas represent the standard error of the mean across subjects. (A) Decoding category and exemplar identity from the four target images presented in the experimental trials. (B) Decoding category and exemplar identity from the 56 images presented in the fast streams at the beginning of each block; category decoding was based on all images in the stream classified by either face or place, and exemplar decoding was based only on the targets and modified targets

## 296 **Significant category decoding in Pattern Estimator**

297 To create a category classification model for imagery, we looked at patterns of brain  
298 activity while participants were viewing images in the fast stream (Pattern Estimator). All  
299 images were labelled according to super-ordinate categories of 'face' or 'place'. To assess  
300 the model's utility, we cross-validated it on the Pattern Estimator trials. There was sustained  
301 above-chance category decoding from 124ms after stimulus onset until approximately  
302 535ms after stimulus onset (Fig 1). The classifier was also able to distinguish between the  
303 two Sydney Harbour Bridge targets at several discrete time points between 236ms and  
304 348ms after stimulus onset. There was no continuous above-chance decoding for Santas.  
305 Category decoding peaked at 404ms after stimulus onset, at 244ms for Sydney Harbour  
306 Bridges, and at 120ms for Santas.

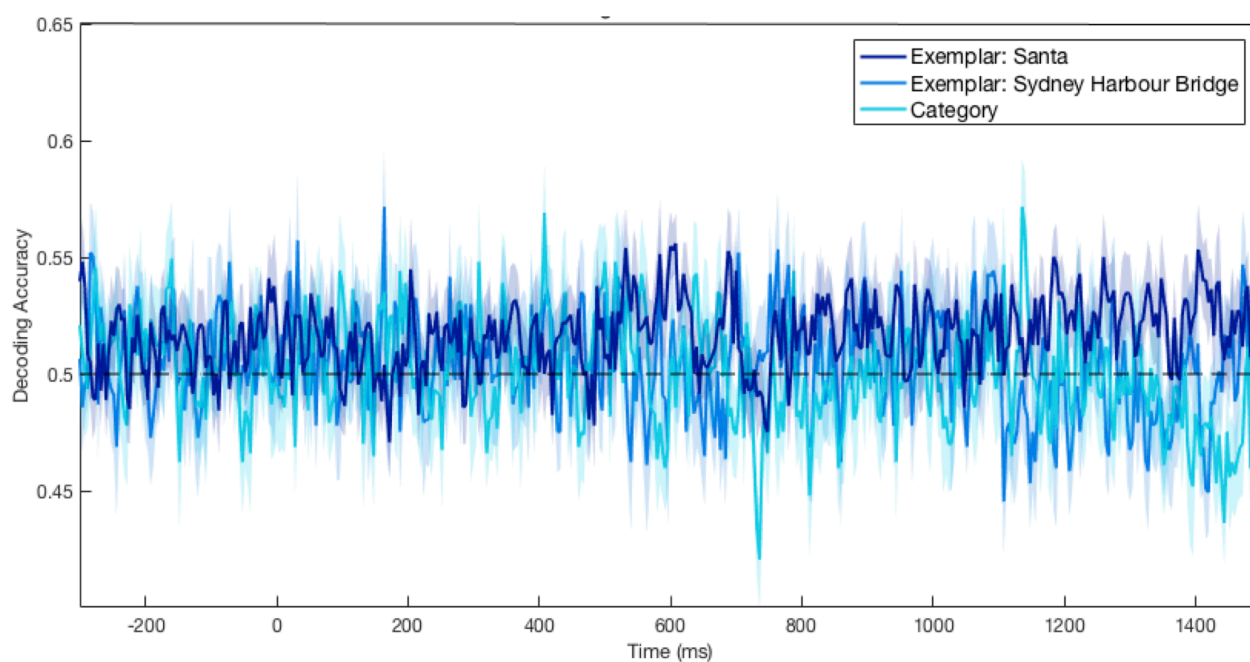
## 307 **No significant decoding for imagery**

308 To determine if category or exemplar information was decodable from imagined data,  
309 we trained and tested a classifier on the Cue- and Response-Locked Imagined epochs (Fig  
310 2). Brain areas activated during imagery are known to vary between individuals (Cui et al.,  
311 2007), so we looked at imagery decoding on an individual subject basis. For each subject,  
312 we ran a permutation test in which the decoding procedure was run 1000 times, with  
313 category labels randomly assigned to the epochs. A  $p$ -value was calculated for each time  
314 point, based on the number of permutations with a greater decoding accuracy than the  
315 correct label decoding. We used the False Discovery Rate to correct for multiple  
316 comparisons. This test was conducted on both Response- and Cue-Locked epochs, and we  
317 found decoding was not significantly above chance for any individual at any time point for  
318 either Cue- or Response-Locked data ( $ps > .05$ ).

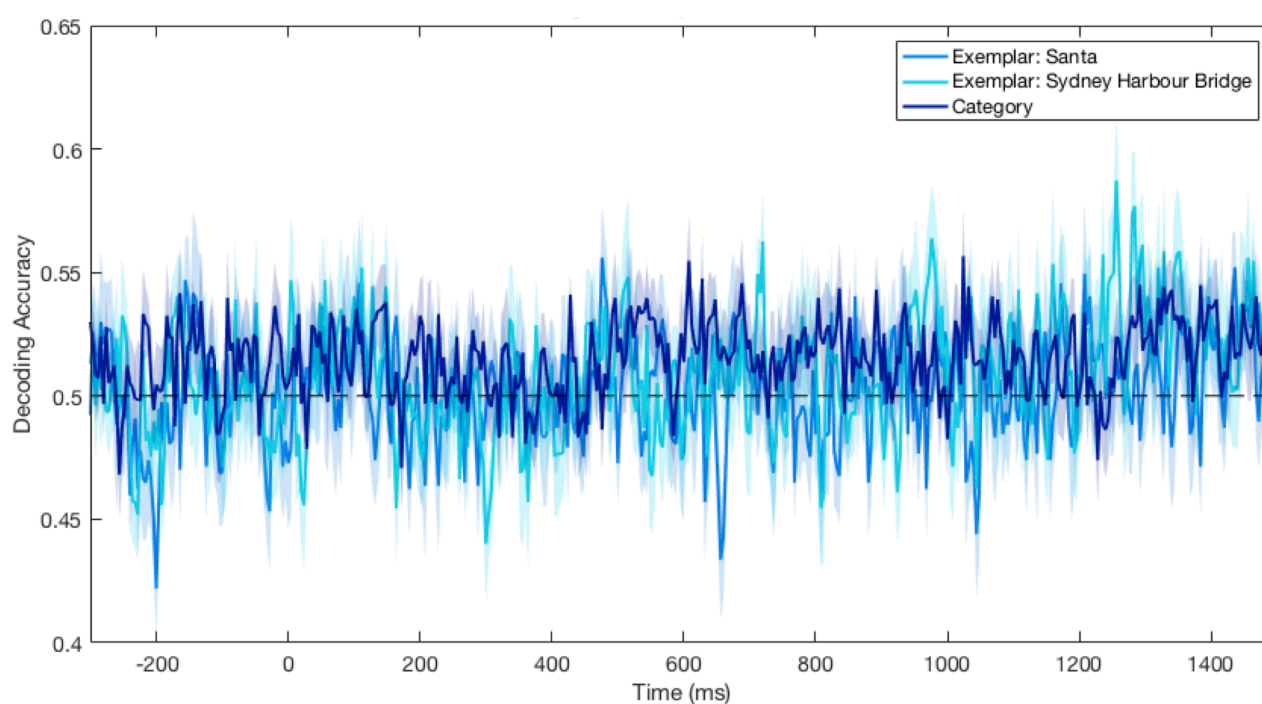
319

320

### (A) Cue-Locked Imagined



### (B) Response-Locked Imagined



**Figure 2.** Mean decoding accuracy for Cue-Locked and Response-Locked Imagined epochs. The absence of dots below the plots indicates there were no points at which decoding was significantly above chance ( $p > .05$ ). Shaded areas represent the standard error of the mean across subjects.

321

322

**(A)** Decoding accuracy centred on when participants click to advance to the imagining period. **(B)** Decoding accuracy centred on presentation of the numerical cue indicating the location of the target in the preceding stream.

323

324 To test whether there was any representational overlap in imagery and vision, we ran a  
325 cross-decoding analysis. We ran all pairwise combinations of vision and imagery; a classifier  
326 trained to distinguish Santas from Harbour Bridges in the viewed stimuli (Pattern Estimator  
327 or Vision epochs) and was tested on imagery periods (Cue-Locked or Response-Locked).  
328 There were no significant periods of overlap for any cross-decoding involving imagined trials  
329 ( $ps > .05$ ).

330 It could be that the processes in vision and imagery engage overlapping representations  
331 but at different times. To test this, we conducted a time generalisation analysis (King and  
332 Dehaene, 2014). A classifier was trained on visual data (Pattern Estimator or Vision epochs)  
333 at each time point, and then tested on imagined data (Cue- and Response-Locked) at every  
334 possible time point. There was no time point where decoding was significantly above chance  
335 for any combination of training and testing (all  $ps > .05$ ), indicating there was no point where  
336 the patterns of brain activity during perceptually processed stimuli were present during  
337 imagery.

### 338 **Differences in vividness did not affect decoding accuracy**

339 Another possibility is that people with greater capacity for imagery have more decodable  
340 imagery representations. To investigate the effects of subjective imagery vividness on  
341 decoding accuracy, we grouped the participants as 'high' or 'low' imagery vividness based  
342 on a median split of their 'eyes-open' scores in the VVIQ. Two participants had the median  
343 score and were excluded from further analysis. We used the eyes-open score because it  
344 was the most relevant for the task at hand, and makes our results comparable to prior MEG  
345 research (Dijkstra et al., 2018), where only the eyes-open section was used. To see if there  
346 were any significant differences between the groups in any of the previously described  
347 analyses, we conducted a random-effects Monte Carlo statistic with 10,000 iterations to find  
348 where differences between the groups were significantly greater than zero. There was only  
349 one isolated point of significant differences between the two conditions, at 1484ms, when



350 the classifier was trained on Pattern Estimator data and tested on Response-Locked  
351 Imagery.

## 352 **Discussion**

353 The current study used time-series decoding to capture the precise temporal  
354 fluctuations underlying mental imagery. Based on prior MEG evidence showing the category  
355 and identity of imagined objects can be decoded, we expected successful category and  
356 exemplar decoding from imagery. However, contrary to our predictions, we were unable to  
357 detect any systematic representations of category or exemplar information during imagery.  
358 Based on previous evidence that imagery recruits similar neural networks to vision (Ganis  
359 et al., 2004), we also anticipated overlapping patterns of neural activity when participants  
360 were viewing and imagining the same image. Although we were able to decode stimulus  
361 category and identity from visually processed stimuli, there were no time points where neural  
362 representations of vision and imagery were overlapping. Finally, we considered whether  
363 individual subject results might vary on the basis of imagery vividness, and found no  
364 systematic differences between subjects reporting high and low vividness. Overall, our  
365 findings demonstrate the variability of imagery processes within subjects over time, and  
366 suggest stimulus- and design-related factors may influence the chances of successfully  
367 decoding mental imagery.

368 To compare the overlap between imagery and visual processing, we first defined  
369 the temporal dynamics of visual processing for the images in this experiment. For stimuli  
370 presented as part of the imagery sequence (Vision), image category was predictable from  
371 approximately 100ms after stimulus presentation until offset 1400ms later. Exemplar  
372 decoding was also significant from 100ms, albeit for less continuous time than category  
373 decoding, reflecting well-established evidence that both categories and exemplars evoke  
374 distinct patterns of brain activity (Carlson et al., 2013). For the Pattern Estimator, category  
375 decoding was significantly higher than chance from 100ms until approximately 500ms after  
376 stimulus onset. This extended period of decoding after stimulus offset supports recent

377 evidence that multiple representations can co-exist in the brain (Grootswagers et al., 2019;  
378 Marti and Dehaene, 2017).

379 In both visual conditions, exemplar decoding peaked earlier than category decoding.  
380 This reflects well-established evidence of increasing abstraction along the ventral visual  
381 pathway (Carlson et al., 2013; Contini et al., 2017). It also appears that decoding accuracy  
382 for Sydney Harbour Bridges is higher than for Santas, for both visual conditions (Vision and  
383 Pattern Estimator), though this pattern is less defined for the Pattern Estimator stimuli  
384 because of the low numbers of training and testing stimuli (4 of each exemplar per stream).  
385 Informal questioning of participants post-experiment suggested many participants found the  
386 Sydney Harbour Bridge images easier to imagine because of the distinct lines forming the  
387 arches and underside of the bridge.

388 When the classifier trained on the visual stimuli was tested on imagery, there were no  
389 time points where the signal was sufficiently similar to accurately predict image category or  
390 identity. To investigate the possibility that the processes were not temporally aligned, we  
391 conducted a temporal generalisation analysis. There were no regular patterns of activity at  
392 the group level, indicating there was no overlap in representations at any point in the imagery  
393 period. Based on evidence that areas of activation during imagery vary across people (e.g.,  
394 Cui et al., 2007), we examined results on the individual level. Patterns of individual decoding  
395 accuracy varied dramatically between subjects. Neither category nor exemplar decoding  
396 was significant at any time point for any individual. At face value, these results seem  
397 inconsistent with prior findings by Dijkstra and colleagues (Dijkstra et al., 2018). These  
398 differences primarily point to the difficulties of studying visual mental imagery, and the  
399 specific methodological characteristics required to obtain significant imagery decoding.

400 Several factors may have impacted our capacity to decode imagined mental  
401 representations. For example, the increased number of channels in MEG compared to EEG  
402 provides better signal to noise ratio and greater likelihood of detecting an effect (Cichy and  
403 Pantazis, 2017). An additional consideration is that individual variability in image generation

404 would reduce the sensitivity of population statistics. Moreover, the temporal variability in an  
405 individual's capacity to generate a mental image would further reduce individual effect sizes.

406 Another potential explanation for our non-significant imagery decoding is the  
407 unavailability of non-imagery based strategies. Previous imagery experiments using a retro-  
408 cue design, in which participants identify the imagery target based on a cue presented  
409 immediately following a sequence of images, have found significant imagery decoding using  
410 only two stimuli (e.g., Dijkstra et al., 2017; Dijkstra et al., 2018; Harrison and Tong, 2009).  
411 However, with only two classes of stimuli, participants can effectively complete the task  
412 without imagery. For example, participants could perform the retro-cue house-face task used  
413 in Dijkstra and colleagues' research (Dijkstra et al., 2018) by recalling a label for each image  
414 as it is presented (e.g., 'house-face'), and mentally repeating this order after cue  
415 presentation. After identifying the target, subjects could simply continue to think of the  
416 relevant label. This pattern of thought is likely to be sufficiently similar during perception and  
417 imagery to be identified by the classifier as a reliable difference between the categories,  
418 leading to accurate decoding of patterns of brain activity based on semantic labels instead  
419 of imagery.

420 We designed our experiment to test if this was the case by including a superordinate  
421 category distinction with two exemplars in each category. We obtained response data after  
422 every trial with flipped images as distractors to test whether participants were using an  
423 imagery-based strategy. If participants were using a purely semantic label-based strategy,  
424 we would expect a similar number of responses for flipped and target images. However, only  
425 0.33% of all responses were the flipped version of the target. These response patterns  
426 clearly show participants in our experiment were aware of the visual elements of the images  
427 rather than solely the semantic label. Due to the fundamentally introspective nature of mental  
428 imagery, there is no way to determine if participants are genuinely completing the imagery  
429 portion of the task. However, these response patterns point strongly to the use of an  
430 imagery-based strategy. Future experiments with similar hierarchical structure and more  
431 subtly modified response options (e.g. deleting or rotating a single element of the image, or

432 changing colours of elements of the target images) could help determine whether this is a  
433 plausible theoretical explanation for our results.

434 Generation of mental imagery requires activation of complex, distributed systems  
435 (Ganis et al., 2004). Higher stimulus complexity increases the number of details that need  
436 to be recalled from memory. It therefore seems likely that the neural processes involved in  
437 viewing a static image are more temporally consistent than generating an image from  
438 memory, which is unlikely to follow a millisecond-aligned time-locked process. This is  
439 particularly apparent for complex stimuli which require more details, stored in potentially  
440 disparate locations, to generate vivid imagery. This same temporal blurring between trials  
441 from temporally misaligned processes is present in other prior studies (Dijkstra et al., 2018),  
442 as it is somewhat inherent to the temporal specificity that decoding of time-series data  
443 provides.

444 Most previous experiments using complex visual scenes as imagery targets use an  
445 extensive training period prior to the study, relying on long-term memories of targets for  
446 imagery (Naselaris et al., 2015). Although our participants completed a training period prior  
447 to EEG recording, slightly longer than those in Dijkstra and colleagues' MEG study, it is  
448 possible (Dijkstra et al., 2018) that participants might have experienced more vivid imagery  
449 if they had more exposure to the experimental images. Intuitively, it seems easier to imagine  
450 a highly familiar object such as an apple rather than a scene of Sydney Harbour because  
451 there are fewer details required to create an accurate representation. Mental images that  
452 are less vivid or less detailed are likely to generate weaker neural activation (Dijkstra et al.,  
453 2017) and are less likely to fully resemble the details that are processed during vision. If the  
454 patterns are less distinct, a classifier is less likely to be able to identify reliable patterns of  
455 brain activity on which to base categorisation. To determine the effects of memory on  
456 imagery vividness and reliability, future study could compare the current results to a similar  
457 paradigm where subjects have extensive training prior to recording (e.g., participants are  
458 extensively questioned about characteristics of the image, or have to draw the main aspects  
459 to show awareness of details in the image).

460 As highlighted in recent research (Dijkstra et al., 2019), individual differences in imagery  
461 generate increased variation between individuals. For example, differences in visual working  
462 memory capacity, personal decision-making boundary, and memory strategy may have  
463 increased variation between participants. Individuals who report stronger imagery ability  
464 tend to use an imagery-based strategy on visual working memory tasks (Pearson et al.,  
465 2015). Features of both working memory and long-term memory (e.g. meaningfulness,  
466 familiarity) affect ratings of imagery vividness (Baddeley and Andrade, 2000). These factors  
467 might also influence variability within a participant... changes over the course of the  
468 experiment, increasing experience with images, etc, could influence temporal variability from  
469 trial-to-trial.

470 Other individual differences, such as personal decision strategies vary across  
471 individuals. We may have captured a slightly different stage of imagery, as it is likely each  
472 person based the timing of their mouse clicks on a different threshold criterion for the point  
473 at which they had begun to imagine. Different strategies for identifying the target may have  
474 directed the focus of imagery. When asked informally at the conclusion of the experiment,  
475 all participants could explicitly describe their strategy for identifying the target. Most  
476 participants assigned a label to each image and mentally repeated these to remember the  
477 image order. The majority of strategies relied on structural characteristics, for example, “fat,  
478 tall, under, above”. Several participants also reported a direction-based strategy, for  
479 example, “top, bottom, centre, side” or “straight, side, face, body”, indicating the orientation  
480 of the main object in the image. Though there is no reliable way to compare decoding  
481 accuracy based on strategy, different strategies may direct focus on different aspects of the  
482 complex images (e.g. thinking of ‘face’ might make facial features salient, compared to  
483 labelling the same image as ‘fat’, drawing focus to body shape). These differences in  
484 strategy present another potential source of variation between subjects.

485 It is clear that capability to decode visual mental imagery is influenced by several  
486 factors, including vividness, memory and stimulus complexity. These factors do not affect  
487 imagery in isolation; they are inherently related. Better memory for the details of an image

488 is likely to increase vividness. The number of details remembered by an individual is  
489 influenced by their memory capacity, but also by the complexity of the stimulus and the  
490 number of details necessary to generate a vivid image. All these factors create variation in  
491 the processes used to generate mental imagery, across both people and time (Borst and  
492 Kosslyn, 2010; Dijkstra et al., 2018). The potential for MVPA techniques to analyse data at  
493 the individual level provides insight into the variation across subjects, and highlights the  
494 need for future studies to consider patterns of data at an individual level to maximise the  
495 chances of obtaining clear signals from imagery.

496

## 497 **Conclusion**

498 In this study, we investigated how neural representations of mental imagery change  
499 over time. Our results suggest successful category decoding in earlier studies may be a  
500 result of better signal to noise ratio from a variety of factors, including individual variation.  
501 Variety in response times, imagery strategy and ability, in addition to fewer recording  
502 sensors may have reduced our power to find systematic patterns of neural activity during  
503 imagery. Furthermore, the interactions between stimulus complexity, working memory, and  
504 imagery vividness may have increased this variation between individuals. Our results raise  
505 many questions for further investigation and demonstrate both the challenges and  
506 advantages associated with time-series decoding for EEG in investigating the introspective  
507 processes underlying mental imagery.

508

## 509 **Supplementary Materials:**

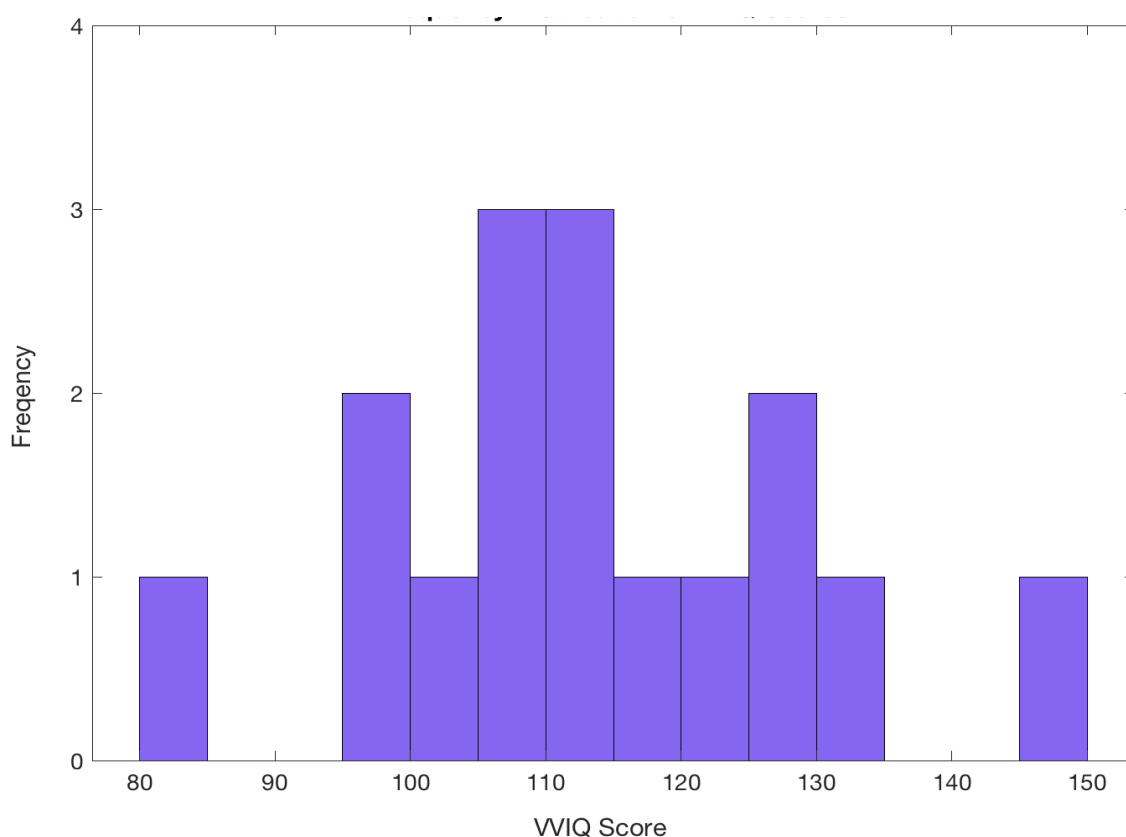
510 **Author Contributions:** Conceptualization, S.M.S.; methodology, S.M.S., T.A.C., A.K.R,  
511 T.G.; formal analysis, S.M.S., T.G.; investigation, S.M.S.; writing—original draft preparation,  
512 S.M.S.; writing—review and editing, S.S., T.A.C., T.G., A.K.R.; supervision, T.A.C., T.G.,  
513 A.K.R.; project administration, T.A.C.; funding acquisition, T.A.C.

514 **Funding:** This research was supported by an Australian Research Council Future  
515 Fellowship (FT120100816) and an Australian Research Council Discovery Project  
516 (DP160101300) awarded to T.A.C. The authors acknowledge the University of Sydney HPC  
517 service for providing High Performance Computing resources. The authors declare no  
518 competing financial interests.

519 **Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role  
520 in the design of the study; in the collection, analyses, or interpretation of data; in the writing  
521 of the manuscript, or in the decision to publish the results.

522

523 **Supplementary data**

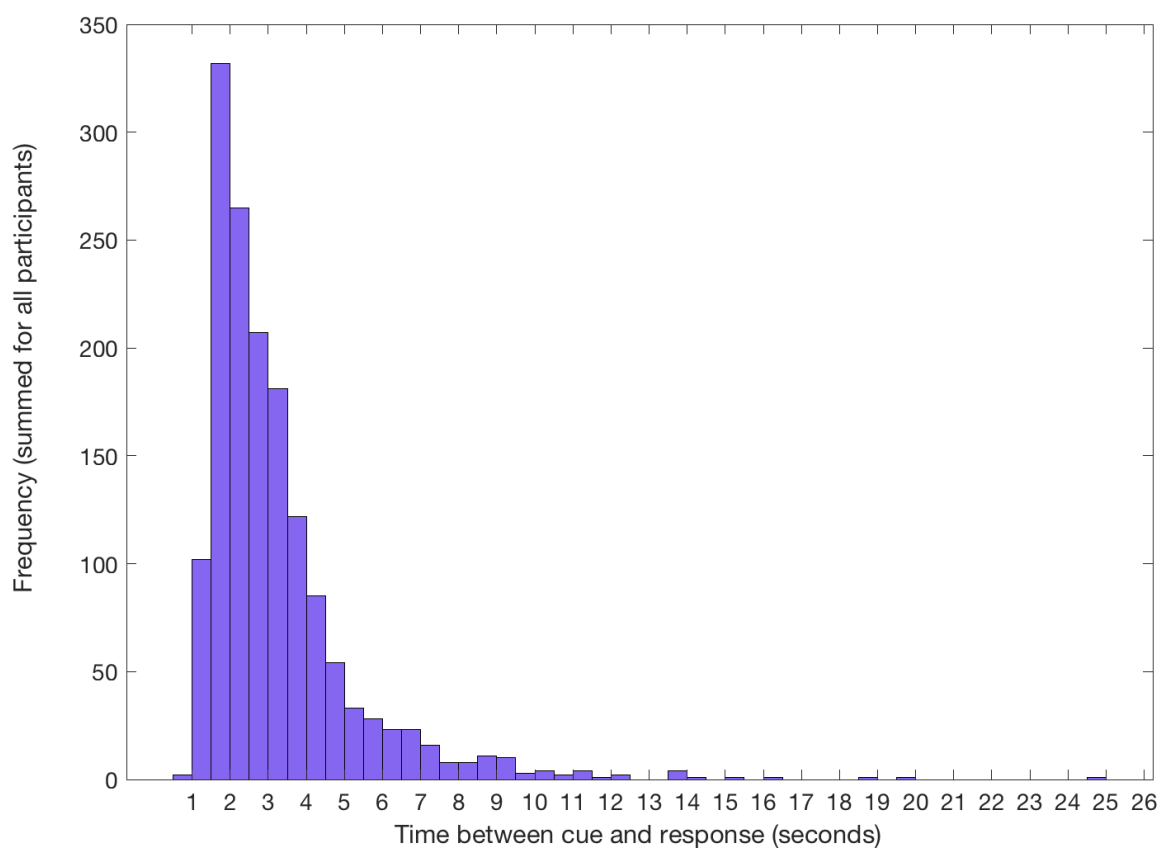


**Figure S1.** Frequency Distribution of scores in the Vividness of Visual Imagery Questionnaire overall scores. Scores are calculated out of a possible 160 by summing responses to each question completed with the eyes open and with the eyes closed.

524

525

526



**Figure S2.** Frequency of response times from cue to imagery across all participants. Response time is taken from the onset of the numerical cue indicating the location of the target in the stream, until the participant voluntarily clicked the mouse. During this period, participants identified the correct target and began to imagine it on the screen.

527

528

529

530

531

532

533



## 534 **References**

- 535 Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., de Lange, F.P., 2013. Shared representations for working  
536 memory and mental imagery in early visual cortex. *Curr Biol* 23, 1427-1431.
- 537 Amedi, A., Malach, R., Pascual-Leone, A., 2005. Negative BOLD differentiates visual imagery and perception.  
538 *Neuron* 48, 859-872.
- 539 Baddeley, A.D., Andrade, J., 2000. Working memory and the vividness of imagery. *J Exp Psychol Gen* 129, 126-  
540 145.
- 541 Bartolomeo, P., Bourgeois, A., Bourlon, C., Migliaccio, R. (Eds.), 2013. *Visual and Motor Mental Imagery After*  
542 *Brain Damage*. Springer New York, New York, NY.
- 543 Borst, G., Kosslyn, S.M., 2010. Individual differences in spatial mental imagery. *Q J Exp Psychol (Hove)* 63,  
544 2031-2050.
- 545 Brainard, D.H., 1997. The Psychophysics Toolbox. *Spat Vis* 10, 433-436.
- 546 Bridge, H., Harrold, S., Holmes, E.A., Stokes, M., Kennard, C., 2012. Vivid visual mental imagery in the  
547 absence of the primary visual cortex. *J Neurol* 259, 1062-1070.
- 548 Carlson, T., Tovar, D.A., Alink, A., Kriegeskorte, N., 2013. Representational dynamics of object vision: the first  
549 1000 ms. *J Vis* 13.
- 550 Chen, W., Kato, T., Zhu, X.H., Ogawa, S., Tank, D.W., Ugurbil, K., 1998. Human primary visual cortex and  
551 lateral geniculate nucleus activation during visual imagery. *Neuroreport* 9, 3669-3674.
- 552 Cichy, R.M., Pantazis, D., 2017. Multivariate pattern analysis of MEG and EEG: A comparison of  
553 representational structure in time and space. *Neuroimage* 158, 441-454.
- 554 Contini, E.W., Wardle, S.G., Carlson, T.A., 2017. Decoding the time-course of object recognition in the human  
555 brain: From visual features to categorical decisions. *Neuropsychologia* 105, 165-176.
- 556 Crawford, H.J., 1982. Hypnotizability, daydreaming styles, imagery vividness, and absorption: a  
557 multidimensional study. *J Pers Soc Psychol* 42, 915-926.
- 558 Cui, X., Jeter, C.B., Yang, D., Montague, P.R., Eagleman, D.M., 2007. Vividness of mental imagery: individual  
559 variability can be measured objectively. *Vision Res* 47, 474-478.
- 560 Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics  
561 including independent component analysis. *J Neurosci Methods* 134, 9-21.
- 562 Dijkstra, N., Bosch, S.E., van Gerven, M.A., 2017. Vividness of Visual Imagery Depends on the Neural Overlap  
563 with Perception in Visual Areas. *J Neurosci* 37, 1367-1373.
- 564 Dijkstra, N., Hinne, M., Bosch, S.E., van Gerven, M.A.J., 2019. Individual differences in the influence of mental  
565 imagery on conscious perception. *bioRxiv* 607770.
- 566 Dijkstra, N., Mostert, P., Lange, F.P., Bosch, S., van Gerven, M.A., 2018. Differential temporal dynamics during  
567 visual imagery and perception. *Elife* 7.
- 568 Fulford, J., Milton, F., Salas, D., Smith, A., Simler, A., Winlove, C., Zeman, A., 2018. The neural correlates of  
569 visual imagery vividness - An fMRI study and literature review. *Cortex* 105, 26-40.
- 570 Ganis, G., Thompson, W.L., Kosslyn, S.M., 2004. Brain areas underlying visual mental imagery and visual  
571 perception: an fMRI study. *Brain Res Cogn Brain Res* 20, 226-241.
- 572 Ganis, G., Thompson, W.L., Mast, F.W., Kosslyn, S.M., 2003. Visual imagery in cerebral visual dysfunction.  
573 *Neurol Clin* 21, 631-646.
- 574 Goldenberg, G., Podreka, I., Steiner, M., Willmes, K., Suess, E., Deecke, L., 1989. Regional cerebral blood flow  
575 patterns in visual imagery. *Neuropsychologia* 27, 641-664.

576 Grootswagers, T., Robinson, A.K., Carlson, T.A., 2019. The representational dynamics of visual objects in rapid  
577 serial visual processing streams. *Neuroimage* 188, 668-679.

578 Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding Dynamic Brain Patterns from Evoked  
579 Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *J Cogn*  
580 *Neurosci* 29, 677-697.

581 Harrison, S.A., Tong, F., 2009. Decoding reveals the contents of visual working memory in early visual areas.  
582 *Nature* 458, 632-635.

583 Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping  
584 representations of faces and objects in ventral temporal cortex. *Science* 293, 2425-2430.

585 Haynes, J.D., 2015. A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives.  
586 *Neuron* 87, 257-270.

587 Ishai, A., Ungerleider, L.G., Martin, A., Haxby, J.V., 2000. The representation of objects in the human occipital  
588 and temporal cortex. *J Cogn Neurosci* 12 Suppl 2, 35-51.

589 Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate  
590 cortex specialized for face perception. *J Neurosci* 17, 4302-4311.

591 King, J.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal  
592 generalization method. *Trends Cogn Sci* 18, 203-210.

593 Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3? *Perception* 36, 14-14.

594 Knauff, M., Kassubek, J., Mulack, T., Greenlee, M.W., 2000. Cortical activation evoked by visual mental  
595 imagery as measured by fMRI. *Neuroreport* 11, 3957-3962.

596 Kosslyn, S.M., Ganis, G., Thompson, W.L., 2001. Neural foundations of imagery. *Nat Rev Neurosci* 2, 635-642.

597 Kosslyn, S.M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J.P., Thompson, W.L., Ganis, G., Sukel,  
598 K.E., Alpert, N.M., 1999. The role of area 17 in visual imagery: convergent evidence from PET and rTMS.  
599 *Science* 284, 167-170.

600 Kosslyn, S.M., Thompson, W.L., 2003. When is early visual cortex activated during visual mental imagery?  
601 *Psychological Bulletin* 129, 723-746.

602 Lee, S.H., Kravitz, D.J., Baker, C.I., 2012. Disentangling visual imagery and perception of real-world objects.  
603 *Neuroimage* 59, 4064-4073.

604 Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related  
605 potentials. *Front Hum Neurosci* 8, 213.

606 Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*  
607 164, 177-190.

608 Marks, D.F., 1973. Visual imagery differences and eye movements in the recall of pictures. *Perception &*  
609 *psychophysics* 14, 407-412.

610 Marti, S., Dehaene, S., 2017. Discrete and continuous mechanisms of temporal selection in rapid visual streams.  
611 *Nat Commun* 8, 1955.

612 Mechelli, A., Price, C.J., Friston, K.J., Ishai, A., 2004. Where bottom-up meets top-down: neuronal interactions  
613 during perception and imagery. *Cereb Cortex* 14, 1256-1265.

614 Moro, V., Berlucchi, G., Lerch, J., Tomaiuolo, F., Aglioti, S.M., 2008. Selective deficit of mental visual imagery  
615 with intact primary visual cortex and visual perception. *Cortex* 44, 109-118.

616 Naselaris, T., Olman, C.A., Stansbury, D.E., Ugurbil, K., Gallant, J.L., 2015. A voxel-wise encoding model for  
617 early visual areas decodes mental images of remembered scenes. *Neuroimage* 105, 215-228.

618 North, B.V., Curtis, D., Sham, P.C., 2002. A note on the calculation of empirical P values from Monte Carlo  
619 procedures. *Am J Hum Genet* 71, 439-441.

620 Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP  
621 measurements. *Clinical Neurophysiology* 112, 713-719.

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

624 Pearson, J., Clifford, C.W., Tong, F., 2008. The functional impact of mental imagery on conscious perception.  
625 *Curr Biol* 18, 982-986.

626 Pearson, J., Naselaris, T., Holmes, E.A., Kosslyn, S.M., 2015. Mental Imagery: Functional Mechanisms and  
627 Clinical Applications. *Trends Cogn Sci* 19, 590-602.

628 Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies.  
629 *Spat Vis* 10, 437-442.

630 Reeder, R.R., Perini, F., Peelen, M.V., 2015. Preparatory Activity in Posterior Temporal Cortex Causally  
631 Contributes to Object Detection in Scenes. *J Cogn Neurosci* 27, 2117-2125.

632 Slotnick, S.D., Thompson, W.L., Kosslyn, S.M., 2005. Visual mental imagery induces retinotopically organized  
633 activation of early visual areas. *Cereb Cortex* 15, 1570-1583.

634 Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing,  
635 threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83-98.

636 Stelzer, J., Chen, Y., Turner, R., 2013. Statistical inference and multiple testing correction in classification-based  
637 multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage* 65, 69-82.

638 Stokes, M., Saraiva, A., Rohenkohl, G., Nobre, A.C., 2011. Imagery for shapes activates position-invariant  
639 representations in human visual cortex. *Neuroimage* 56, 1540-1545.

640 Zago, S., Corti, S., Bersano, A., Baron, P., Conti, G., Ballabio, E., Lanfranconi, S., Cinnante, C., Costa, A.,  
641 Cappellari, A., Bresolin, N., 2010. A cortically blind patient with preserved visual imagery. *Cogn Behav Neurol*  
642 23, 44-48.

643