The Journal of Neuroscience

https://jneurosci.msubmit.net

JN-RM-0766-23R1

Suppressive Control of Incentive Salience in Real-World Human Vision

Clayton Hickey, Universita degli Studi di Trento David Acunzo, University of Birmingham Jaclyn Dell, University of Birmingham

Commercial Interest:

This is a confidential document and must not be discussed with others, forwarded in any form, or posted on websites without the express written consent of The Journal for Neuroscience.

1	
2	Suppressive Control of Incentive Salience in Real-World Human Vision
3	
4	Clayton Hickey, David Acunzo, & Jaclyn Dell
5	Centre for Human Brain Health and School of Psychology
6	University of Birmingham, UK
7	
8	
9	Correspondence to:
10	c.m.hickey@bham.ac.uk
11	
12	CH designed the experiment; DA and JD collected data; CH analysed the data; CH wrote the
13	paper.
14	
15	This work was supported by the European Research Council (ERC) under the European Union's
16	Horizon 2020 research and innovation programme (grant agreement 804360 to CH). Thanks to
17	Daniele Pollicino and Giacamo Bertazzoli for stimuli preparation and Damiano Grignolio for
18	discussion.
19	

21 Abstract

22 Reward-related activity in the dopaminergic midbrain is thought to guide animal behaviour, in 23 part by boosting the perceptual and attentional processing of reward-predictive environmental 24 stimuli. In line with this incentive salience hypothesis, studies of human visual search have 25 shown that simple synthetic stimuli – like lines, shapes, or Gabor patches – capture attention to 26 their location when they are characterized by reward-associated visual features like colour. In 27 the real world, however, we commonly search for members of a category of visually-28 heterogenous objects - like people, cars, or trees - where category examples do not share low-29 level features. Is attention captured to examples of a reward-associated real-world object 30 category? Here, we have human participants search for targets in photographs of city- and 31 landscapes that contain task-irrelevant examples of a reward-associated category. We use the 32 temporal precision of EEG machine learning and ERPs to show that these distractors acquire 33 incentive salience and draw attention, but do not capture it. Instead, we find evidence of rapid, 34 stimulus-triggered attentional suppression, such that the neural encoding of these objects is 35 degraded relative to neutral objects. Humans appear able to suppress the incentive salience of 36 reward-associated objects when they know these objects will be irrelevant, supporting the rapid 37 deployment of attention to other objects that might be more useful. Incentive salience is thought 38 to underlie key behaviours in eating disorders and addiction, among other conditions, and the 39 kind of suppression identified here likely plays a role in mediating the attentional biases that 40 emerge in these circumstances.

42 Significance Statement

43 Like other animals, humans are prone to notice and interact with environmental objects that 44 have proven rewarding in earlier experience. However, it is common that such objects have no 45 immediate strategic use and are therefore distracting. Do these reward-associated real-world 46 objects capture our attention, despite our strategic efforts otherwise? Or are we able to 47 strategically control the impulse to notice them? Here we use machine learning classification of 48 human electrical brain activity to show that we can establish strategic control over the salience 49 of naturalistic reward-associated objects. These objects draw our attention, but do not 50 necessarily capture it, and this kind of control may play an important role in mediating conditions 51 like eating disorder and addiction.

53 Introduction

54 Humans and other animals preferentially approach stimuli that have been associated with 55 positive outcome in prior experience, and this is thought to involve an impact of reward on 56 perception and attention. By this incentive salience hypothesis, reward-elicited activity in the 57 dopaminergic midbrain impacts perceptual systems, causing reward-predictive stimuli to 58 become salient and attention-drawing and ensuring the information carried by these objects 59 gains access to decision making and motor control (Berridge & Robinson, 1998). This bias is 60 thought to be independent of strategy, with reward-associated stimuli drawing attention even 61 when this is inconsistent with goals.

62 In line with this, visual search experiments in humans have shown that irrelevant reward-63 associated stimuli interfere with task-focussed behaviour (eq. Della Libera & Chelazzi, 2009; 64 Hickey, Chelazzi, & Theeuwes, 2009; Anderson, Laurent, & Yantis, 2011; Le Pelley, Pearson, 65 Griffiths, & Beesley, 2015) and this has been linked to activity in dopaminergic brain nuclei 66 (Hickey & Peelen, 2015, 2017; Barbaro, Peelen, & Hickey, 2017) and to the concentration of 67 intrasynaptic dopamine in these areas (Anderson et al., 2016). The representative behavioural 68 finding is that responses to a target are slower and less accurate when the environment 69 contains a reward-associated distractor. Though this behavioural effect is ambiguous - it is 70 potentially a product of filtering costs and the need for cognitive control rather than the capture 71 of attention – results from EEG and MEG have convincingly demonstrated that attention is 72 deployed to the reward-associated stimulus, for example by showing that reward-associated 73 distractors elicit an N2pc (Luck & Hillyard, 1994), a component of the event-related potential 74 (ERP) linked to attentional selection and resolution (eq. Hickey, Chelazzi, & Theewues, 2009; 75 Qi, Zeng, Ding, & Li, 2013; Donohue, Hopf, Bartsch, Schoenfeld, Heinze, & Woldorff, 2016). 76 Similarly, MRI results have demonstrated sensitivity to reward-associated distractors in early 77 visual cortex (Itthipuripat, Vo, Sprague, & Serences, 2019).

78 Importantly, this existing body of work has relied on visual search arrays composed of 79 synthetic objects - circles, squares, lines or Gabor patches presented in regular arrays and 80 characterized by saturated primary colours. In roughly the same timeframe as these studies, a 81 separate literature has demonstrated that the exclusive use of such stimuli can lead to 82 misunderstanding of the mechanisms that support visual search (Peelen & Kastner, 2014, for 83 review). Naturalistic search through real-world images is faster than work with synthetic stimuli 84 has suggested should be the case (Thorpe, Fize, & Marlot, 1996), possibly due to the 85 constraining influence of scene semantics and gist (Torralba, Oliva, Castelhano, & Henderson, 86 2006; Wolfe, Võ, Evans, & Greene, 2011), and real-world search is sensitive to issues like 87 target and distractor familiarity (Hershler & Hochstein, 2009; Mruczek & Sheinberg, 2005) and 88 the characteristic positioning of objects in a scene (Kaiser, Quek, Cichy, & Peelen, 2019).

89 This motivates the need for dedicated investigation of naturalistic incentive salience. 90 Results from experiments with scene stimuli demonstrate that examples of reward-associated 91 real-world object categories disrupt behavioural responses to targets (Hickey, Kaiser, & Peelen, 92 2015). Multivoxel classification analysis of fMRI has shown that ventral visual cortex carries 93 more information about a naturalistic reward-associated target than it does a neutral target, but 94 less information about a reward-associated distractor, and this has been interpreted as 95 evidence of the misallocation of attention (Hickey & Peelen, 2015, 2017; Barbaro, Peelen, & 96 Hickey, 2017). The idea here is that the fleeting capture of attention to the distractor is not 97 reflected in the hemodynamic fMRI signal because of the low temporal accuracy of this 98 measure. Instead, fMRI indexes the long-lived post-capture suppression of the distractor that 99 allows attention to be redeployed in search for the target.

100 These behavioural and imaging results have therefore been interpreted as evidence that 101 reward-associated naturalistic distractors capture attention, but this clearly rests on a pair of 102 questionable assumptions. The first is that the behavioural cost of a reward-associated 103 distractor in naturalistic search necessarily reflects its ability to capture attention; as noted



Figure 1 – Examples of scene stimuli. The two left columns contain examples of trees; the two right columns contain examples of bushes. The rows are organized according to the presence and position of people and cars.
 107

109 above, the alternative is that these objects are not selected, but nevertheless degrade 110 behaviour, for example by creating the need for attentional filtering or cognitive control (Folk & 111 Remington, 1998; Sawaki & Luck, 2010; Gaspelin & Luck, 2019). The second assumption is 112 that the suppression of reward-associated distractors observed in fMRI is a reaction to 113 preceding attentional selection; the alternative is that the reward-associated distractor is 114 suppressed from its first appearance. In the current study, we leverage the temporal precision of 115 EEG machine-learning classification and ERPs to directly test the idea that attention is captured 116 to examples of reward-associated distractor categories presented in photographs of real-world 117 scenes.

118

119 Materials and Methods

120 We had participants search through photographs of scenes for examples of real-world 121 categories – cars, people, and plants – and report a characteristic of the target category (Figure 122 1). When the target category was cars or people, participants reported the facing direction of the 123 target; when the target was plants, they reported if the scene contained trees or bushes. 124 For each participant, a single category – always either cars or people – was associated with reward. In blocks where this category was the cued target, correct performance earned 100 125 126 points with cash value (Figure 2a). When any other category was the cued target, correct 127 performance earned only 1 point (Figure 2b). Critically, when participants were cued to search 128 for a low-reward target category, the scene could contain an example of the high-reward 129 category as a task irrelevant distractor (Figure 3). Our core interest lay in these conditions, 130 where we could isolate the neural response to an example of a non-target object as a function 131 of its prior reward association. 132

133



Figure 2 – Trial examples when search is for 'cars'. A.) In subject group A the high-reward category is 'cars', and in this example the current target of search is 'cars'. The scene contains an example of the target category; for this subject group this is a 'lateral high-reward target'. The task is to report the facing direction of the target, which is left, and correct response garners high-magnitude reward. B.) In subject group B the high-reward category is 'people', but in this example the current target of search is 'cars'. The scene contains an example of the target category is 'people', but in this example the current target of search is 'cars'. The scene contains an example of the target category; for this subject group this is a 'lateral low-reward target'. The task is to report the facing direction of the target, which is left, and correct response garners low-magnitude reward.

144

145 The scene stimuli employed in the experiment are visually heterogenous, and physical 146 differences in the images and categories could drive variance in the neural response that might 147 obscure effects of reward association. The experiment had two key features to control for this. 148 First, there were two groups of participants: one group associated reward with cars, the other 149 with people. Second, plants were never associated with reward. Critical conditions were 150 therefore those where participant groups were cued to search for the same neutral low-reward 151 category - plants - and the scenes additionally contained a lateral example of either the high-152 reward distractor category or the low-reward distractor category (or both, with one presented 153 laterally and the other centrally; see Figures 1 and 3).



- 155 Figure 3 – Trial examples for the critical conditions when search is for the low-reward category 'plants'. In group A the 156 high-reward category is 'cars', while in group B the high-reward category is 'people'. The scenes can contain 157 examples of people or cars, or examples of both categories concurrently. When people and cars are both present in 158 the scene, one example is presented laterally and the other centrally. A.) The scene contains an example of the non-159 target category people. For group A, this scene contains a 'lateral low-reward distractor', whereas for group B this 160 same scene contains a 'lateral high-reward distractor.' B.) The scene contains an example of the non-target category 161 cars. For group A, this scene contains a 'lateral high-reward distractor', whereas for group B this same scene 162 contains a 'lateral low-reward distractor'. C.) The scene contains a lateral example of the non-target category people 163 and a central example of the non-target category cars. For group A, this scene contains a 'lateral low-reward 164 distractor' and a 'central high-reward distractor'. For group B this same scene contains a 'lateral high-reward distractor' and a 'central low-reward distractor'. D.) The scene contains a lateral example of the non-target category 165 166 cars and a central example of the non-target category people. For group A, this scene contains a 'lateral high-reward 167 distractor' and a 'central low-reward distractor'. For group B, this scene contains a 'lateral low-reward distractor' and a 168 'central high-reward distractor.' 169
- 170

171 Critically, by manipulating the reward association across participant groups, we were able to use the same scenes in each of these experimental conditions. As an example of this, 172 173 consider the trial illustrated in Figure 3a. For group A, the high-reward target category is cars, 174 but in this example the current target category is plants. The scene contains a single example of 175 a person (alongside examples of the target). For participants in group A, this scene therefore 176 contains a lateral example of the low-reward distractor category. However, for participants in 177 group B, this same scene contains a lateral example of the high-reward distractor category. 178 When results were collapsed across groups, physical differences in the scene stimuli were

179 counterbalanced across participant groups.

In analysis, we use machine learning of EEG data to measure the quality of encoding and representation of reward-associated and neutral distractors, subsequently unpacking classification results through consideration of ERPs. In ERP analysis, our focus lay particularly on the N2pc (Luck & Hillyard, 1994) and Pd components (Hickey, Di Lollo, & McDonald, 2009) as indices of attentional selection and suppression, respectively. To foreshadow, classification and ERP results demonstrate that naturalistic reward-associated distractors are strongly suppressed from the moment they appear.

187

188 Participants

Thirty-six healthy volunteers from the University of Birmingham community gave informed consent before completing the experiment. Each participant reported normal or corrected-to-normal vision and was paid for participation. Two participants were rejected from analysis due to poor accuracy in low-reward task conditions (>2.5 SD from the mean), leading to a final sample of 34. Of these, 3 were left-handed, 10 were men, and mean age was 20 years (3 years SD).

195

196

197 Stimuli and Procedure

198 Participants searched through black and white photographs of real-world scenes 199 (approximately 22° x 17° visual angle) for examples of three different object categories. Figure 1 200 presents a set of scene examples. The target category changed for each of the 24 experimental 201 blocks, with each block containing 54 trials, and a cue at the beginning of each block identified 202 whether cars, people, or plants were the target category for that set of trials. When participants 203 were cued to search for cars or people, every scene in the block contained a single example of 204 the target category located at the left, middle, or right of the image. Participants were asked to 205 report the facing direction of the target – for example, if the car faced the left or the right – via 206 button press with the left or right hand on a standard computer keyboard. When participants 207 were cued to search for plants, the scenes contained at least one example of a tree or a bush, 208 but not examples of both, and these were located anywhere in the scene. Participants reported 209 whether the scene contained trees or bushes with a corresponding left- or right-hand keyboard 210 response. The target category for each individual block was selected at random with the 211 constraint that each category served as target for an equal number of blocks in the experiment. 212 The scenes could contain examples of the categories not currently acting as target. For 213 example, when search was for people, the scene could contain examples of cars and plants as 214 task-irrelevant nontargets. When these distractors were cars or people, only a single 215 example appeared and was located at the left, middle, or right of the scene. When the 216 distractors were plants, multiple examples could appear at any location. 217 When people or cars were the target category, scenes were constrained such that they

contained either a lateral example of the target category, a lateral example of the target
category and a central example of the other localized distractor category, or a lateral example of
the localized distractor category and a central example of the target category (see Figure 1).
Equal numbers of these *target lateral, target lateral / distractor central*, and *distractor lateral /
<i>target central* scenes were presented. In each of these layouts, the facing direction of the target

and the facing direction of the localized distractor were counterbalanced across images withineach category, as was the presence of trees or bushes.

When plants were the target category, scenes were constrained such that they contained either a lateral example of a car distractor, a lateral example of a car distractor and a central example of a person distractor, a lateral example of a person distractor, or a lateral example of a person distractor and a central example of a car distractor. Equal numbers of these *lateral distractor* and *lateral distractor / central distractor* scenes were presented to the participant and the facing direction of the distractors was counterbalanced across images within each category.

232 There were 304 scene images in the stimuli set, most taken from a set of 480 images 233 employed in an earlier publication (Hickey, Pollicino, Bertazolli, & Barbaro, 2019) with additional 234 scenes generated using a digital camera. Each core image set (eg. left-located left-facing 235 car, central right-facing person, bush) had 4 to 8 individual examples. Examples from each core 236 image set were used in the experiment in random order until all images in the set had been 237 presented, at which point this process reset in new random order. The scene images were 238 prepared such that the category example in the periphery was roughly equidistant from fixation 239 in each image and such that people and vehicles had roughly consistent size across the image 240 set.

In each trial, correct response was rewarded with points that had cash value, with the magnitude of reward varying as a function of target category. For 17 of the participants, correct responses to car targets resulted in high-magnitude reward (100 points), whereas correct responses to people or plant targets resulted in low-magnitude reward (1 point). For the remaining participants, correct responses to people resulted in high-magnitude reward, with cars and plants associated with low-magnitude reward. The points putatively determined a final pay range of £18 to £24, and participants were instructed to maximize points and therefore

earnings, but at the end of the experiment total earnings were rounded to £24 for allparticipants.

250 The experiment took place in a dimly lit room and participants were seated at 251 approximately 1 m distance from a 24-inch LED monitor (100 Hz refresh rate). As illustrated in 252 Figure 2, each experimental block began with a 2 s cue indicating the target category for that 253 block and a reminder of the high-reward category. Each trial began with presentation of a 254 fixation cross for 250 to 750 ms (randomly selected from a uniform distribution) followed by 255 presentation of a scene for 200 ms. The scene was subsequently replaced by a fixation cross 256 until either the participant responded via keyboard button press or 1750 ms had passed. 257 Reward feedback was then presented for 1000 ms, after which a new trial began. Feedback 258 regarding task accuracy and speed was provided at the end of every experimental block and the 259 session took approximately 2.5 hours, reflecting 1.5 hours of experimental participation and 1 260 hour of preparation and debriefing. Stimuli presentation relied on PsychToolbox-3 for MATLAB 261 (Brainard et al., 2008).

262

263 EEG Recording and Pre-processing

264 EEG was recorded at 1 kHz from 64 Ag/AgCl electrodes mounted in an elastic cap using a 265 Biosemi Active2 amplifier and ActiView acquisition software. Horizontal electrooculogram (EOG) 266 was recorded from electrodes 1 cm lateral the left and right external canthi, vertical EOG was 267 recorded from electrodes place directly above and below the left pupil, and two additional 268 electrodes recorded voltage over the left and right mastoid processes. Electrode offset was 269 minimized and stabilized prior to the start of recording. EEG was acquired at DC with a 208 Hz 270 anti-aliasing filter, resampled offline at 512 Hz, re-referenced to the average of mastoid signals, 271 and band-pass filtered with a Hamming windowed FIR kernel (0.1 to 45 Hz; -6dB at 0.05 Hz and 272 45.05 Hz). Epochs beginning 1 s before and ending 2 s after each scene onset were extracted 273 from the data.



²⁷⁶

277 Figure 4 – Behavioural results from conditions where 'cars' or 'people' were the target of search. These results are 278 presented for the sake of completeness; no core hypotheses are tested. In the stylized scene examples employed 279 here and in subsequent figures high-magnitude reward is associated to 'cars', but as described in the body of the 280 paper this was counter-balanced across participants. A.) Accuracy. As expected, responses to high-reward targets 281 were more accurate in all conditions. B.) Reaction times. Surprisingly, participants were faster to respond to targets 282 presented in scenes that also contained an example of the localized distractor. This may reflect a qualitative 283 284 difference in the images; scenes containing only one localized category type happened to be characterized by smaller, harder-to-find target examples. Data collected from presentation of these scenes were not employed to test 285 the core study hypothesis regarding the capture of attention to reward-associated stimuli. 286

287	Infomax independent component analysis (Bell & Sejnowkski, 1995) was used to identify
288	variance stemming from ocular artifacts in the epoched data. The independent components
289	representing horizontal and vertical eye movements were used to identify trials in which eye
290	movements were made in the 600 ms interval following stimulus onset. Participants moved their
291	eyes in 6 – 18% of trials and these were removed from further analysis. Components
292	representing eye and muscle artifacts were subsequently removed from the data, as were trials
293	resulting in incorrect response, and epoched data was baselined on the 200 ms interval
294	preceding scene onset. Experimental conditions were subsequently defined based on the



296

297 Figure 5 – Results from conditions where participants searched for plants, which were employed to test motivating 298 experimental hypotheses regarding attentional capture. A.) When participants searched for plants in scenes 299 containing a single additional distractor, accuracy degraded when that distractor was associated with reward. 300 However, when the scene contained both a lateral and a central distractor - and therefore always contained 301 examples of both the high-reward and low-reward distractor categories - accuracy was insensitive to the specific 302 locations of the 2 distractors. C.) Similar results emerge in reaction times. When participants searched for plants in 303 scenes containing a single additional distractor, reaction times increased when that distractor was associated with 304 reward. However, when the scene contained 2 distractors, reaction time was insensitive to the specific location of 305 these distractors.

306

307 reward association of the target category (reward-associated car / person, neutral car / person,

- 308 or neutral plant) and the presence and location of distractor stimuli.
- 309

310 EEG machine learning classification

- 311 Our approach to EEG classification is based on linear discriminant analysis (LDA) and
- 312 cross-fold validation. Each classification analysis interrogates a conditional difference, for
- 313 example whether a target is located on the left or right of the scene, with the classifier trained to
- 314 label data as coming from one of these two classes. In each analysis, conditional EEG is
- 315 partitioned into 10 folds, each balanced to contain an equal number of randomly selected,
- 316 correctly performed trials from each of the two classes, and a model is built for each



318 Figure 6 - Results from trial-wise EEG classification. In all analyses, the machine learning algorithm is trained to 319 discriminate between two classes of stimuli and 50% accuracy reflects chance performance. Topographic maps 320 321 reflect model decision criteria across reward conditions as measured over the latency intervals identified by grey shading in the corresponding time-course plots, which is a 40 ms interval centred on peak decoding accuracy 322 collapsed across conditions. Topographic plots are computed as the mean of spatially z-scored data for each 323 individual and are therefore in SD units; while the topographical pattern is informative, the underlying values are 324 uninformative and as such no scale is provided in the figure. Significant classifier accuracy in each condition is 325 illustrated in the red or blue lines located just above the 50% baseline. a.) Results from classification of target 326 location. This analysis is intended to demonstrate the efficacy of the method, and to investigate the impact of reward 327 association on target processing, but does not test the motivating experimental hypothesis regarding the capture of 328 attention. Analysis does not identify a reliable difference in classification accuracy as a function of target reward 329 association. b.) Results from classification of distractor location. Location classification improves when the distractor 330 is associated with reward. c.) Results from classification of whether the distractor was present in the centre of the 331 scene. Presence classification degrades when the distractor is associated with reward. 332

333 combination of 9 data folds. The 10 resulting models - each based on a unique combination of 334 9 of 10 data folds – are subsequently tested against the individual trials contained in the single 335 fold that did not contribute to model building. There is no trial averaging in our approach, and 336 classification accuracy is defined as the mean testing accuracy across trials and folds. To 337 establish a time-course of classification accuracy, we implemented this modelling and validation 338 procedure for each ~2ms sample point in an epoch beginning 250 ms before the onset of the 339 scene stimulus and ending 1000 ms after. To ensure model stability and accuracy, models were 340 built and tested on data spanning a 61-sample interval centred on the datapoint under 341 consideration (constituting 64 * 61 = 3904 observations of electrode voltage). Each datapoint in 342 classification analyses thus reflects classification performance across a ~120 ms interval 343 centred on the datapoint under consideration. This importantly means that the absolute latency 344 of classification onset should be interpreted with care, as accuracy at a given timepoint reflects 345 the performance of a model with access to data recorded up to ~60 ms later. In contrast, peak 346 classification latency and conditional effects on classification latency can be unambiguously 347 interpreted.

348 To gain insight on model classification decisions, we extracted model weights in each 349 model building instance. These were subsequently multiplied by the covariance matrix of the 350 data that had been used to build the model, with the results mean averaged across model 351 building iterations and across latency intervals of interest and z- scored within each participant 352 before being mean averaged across participants. This procedure transforms the backward 353 model generated by LDA, which projects a data pattern into an expected class membership, into 354 a forward model, which projects class membership into an expected data pattern (Haufe et al, 355 2014). The forward model can be topographically plotted to illustrate the classifier decision 356 criteria (see Figure 6).

357 Statistical analysis of classification accuracy relied on threshold-free cluster 358 enhancement (TFCE; Smith & Nichols, 2009) with clusters defined over time. Conditional 359 differences in classification accuracy were tested using permutation contrasts with 100k 360 iterations based on mean accuracy observed in a 40 ms interval centred on the cross-361 conditional accuracy peak. Statistical analysis of the latency of classification accuracy relied on 362 a resampling approach. To assess the difference in peak classification latency between 363 conditions we iteratively resampled from the set of 34 participant datasets 100k times with 364 replacement. In each iteration, we averaged classification accuracy for the relevant conditions 365 across the sample, extracted the peak latency for each condition, and calculated the difference 366 in peak latencies. The probability that an observed difference in peak classification latency 367 might have been observed under the null hypothesis was reflected in the proportion of the 368 distribution of difference scores that fell below zero. Classification analyses relied on the 369 COSMOMVPA (Oosterhof, Connolly, & Haxby, 2016) and ADAM toolboxes (Fahrenfort, Van 370 Driel, Van Gaal, & Olivers, 2018) and on custom code.

371

372 Event-related potentials

373 ERPs were calculated using standard signal-averaging (Luck, 2014). Our focus 374 was on the N2pc and Pd components of the visual ERP, which index attentional selection and 375 attentional suppression, respectively, and emerge in visual cortex contralateral to the location of 376 the eliciting stimulus. To isolate these components from bilateral variance in the ERP, we a.) 377 extracted voltage recorded at electrodes located over left visual cortex when the eliciting 378 stimulus was in the right visual field, and averaged this response with b.) voltage recorded at 379 electrodes located over right visual cortex when the eliciting stimulus was in the left visual field. 380 This generated a contralateral waveform and a similar procedure was applied to generate 381 ipsilateral waveforms. Topographic maps of differences in lateralized ERP components are 382 generated by 'flipping' EEG data observed when the eliciting stimulus is in the right visual field

384

385

and averaging with EEG data observed when it is in the left visual field, such that the left cortical
 hemisphere consistently represents ipsilateral cortex and the right cortical hemisphere
 consistently represents contralateral cortex.

389 Importantly, when calculated in reference to the objects appearing to the left and right of 390 fixation, the N2pc and Pd are insensitive to lateralized activity evoked by objects in the center of 391 the visual field (Woodman & Luck, 2003; Hickey, McDonald, & Theeuwes, 2006; Hickey, Di 392 Lollo, & McDonald, 2009). For example, consider a display with a central car distractor and a 393 lateral person target, with the central distractor eliciting theoretical right-lateralized ERP activity. 394 When the person target is in the left visual field, the car-elicited effect emerges as positivity in 395 the contralateral signal. But when the person distractor is in the right visual field, the car-elicited 396 effect emerges as negativity in the contralateral signal. When mean target-elicited contralateral 397 signal is calculated, the central distractor has no summed effect. 398 Statistical analysis of ERP component amplitude depended on parametric repeated-399 measures ANOVA (RANOVA). Lateral ERPs were statistically assessed in two latency intervals: 400 220 – 280 ms, when the N2pc and Pd are known to emerge with maximum amplitude (Luck & 401 Hillyard, 1994; Hickey, Di Lollo, & McDonald, 2009), and 100 – 160 ms, when an early 402 expression of the Pd is known to emerge (Weaver, van Zoest, & Hickey, 2017; Sawaki & Luck, 403 2010). ERP analysis relied on the EEGLAB toolbox (Delorme & Makeig, 2004) and custom 404 code. Additional control analyses involving linear mixed models and Bayesian model 405 comparison are described in the results section and depend on the *fitlme.m* function 406 implemented in the MATLAB statistic toolbox (R2021b) and the BayesFactor toolbox 407 (https://klabhub.github.io/bayesFactor) with default priors.



Figure 7 – Relationship between classification of distractor location and distractor presence across participants.
 Statistical analysis reflects permutation analysis with 100k iterations, an approach that is robust to the influence of outliers.

413

424

- 414 Results
- 415 Behaviour

416 Outliers were defined as responses where reaction time (RT) was more than 3 SD from 417 the participant mean and were rejected from further analysis (1.3% of trials, 0.4% SD). Accuracy 418 and RT are illustrated in Figures 4 and 5. The results illustrated in Figures 4a and 4b are 419 presented largely for the sake of completeness as no critical experimental hypotheses are 420 tested. As expected, participants were more accurate when correct response to the target 421 garnered high-magnitude reward (Figure 4a). Unexpectedly, they were slower to respond to the 422 target when the distractor was absent from the scene (Figure 4b). 423 Results from the critical plant target condition are presented in Figure 5. When a lateral

distractor was present in the scene during search for a plant, the association of reward to the

distractor decreased accuracy (Figure 5a, left of panel) and increased RT (Figure 5b, left of
panel). When the scene contained both a lateral distractor and a central distractor – and
therefore contained both a reward-associated and a neutral distractor – the specific location of
the reward-associated and neutral distractors did not have an impact on accuracy (Figure 5a,
right of panel) or RT (Figure 5b, right of panel).

430 Statistical analysis of accuracy in the plant target condition took the form of a RANOVA 431 with factors for group (cars are reward-associated vs. people are reward-associated), reward 432 association of lateralized distractor (reward-associated lateral distractor vs. neutral 433 lateral distractor), and presence of central distractor (central distractor present vs. central 434 distractor absent). This identified a main effect of distractor reward association (F(1,32) = 18.68, 435 p < 0.001, η_p^2 = 0.369), reflecting the decrease in accuracy when a reward-associated distractor 436 was present, and an interaction of distractor reward association and central distractor presence 437 (F(1,32) = 22.19, p < 0.001, η_p^2 = 0.410), reflecting accentuation of this effect when the central 438 distractor was absent from the scene. No other effects reached significance (central distractor 439 presence: F(1,32) = 3.09, p = 0.088; group X distractor reward association: F(1,32) = 2.91, p = 440 0.098; group X central distractor presence: F(1,32) = 2.66, p = 0.113; 3-way interaction: F(1,32)441 = 1.74, p = 0.196; all other Fs < 1).

A similar pattern of results emerged from analysis of RT. A RANOVA with the same factors identified a main effect of distractor reward association (F(1,32) = 41.30, p < 0.001, η^{2}_{p} = 0.563) and a main effect of central distractor presence (F(1,32) = 94.84, p < 0.001, η^{2}_{p} = 0.747), alongside an interaction of distractor reward association and central distractor presence (F(1,32) = 59.12, p < 0.001, η^{2}_{p} = 0.649). No other effects reached significance (group: F(1,32) = 1.73, p = 0.198; all other Fs < 1).

448

The critical observation from these behavioural results is that the reward-associated

distractor decreased accuracy and increased RTs relative to the neutral distractor. As we found
no behavioural differences between the two subject groups, we collapse across this distinction
in subsequent analysis of electrophysiological data.

452

453 EEG classification

454 As illustrated in Figure 6, we conducted three independent classification analyses of 455 EEG data. The first was focussed on trials where participants searched for cars or people, 456 separating these into conditions based on the location of the target (left hemifield vs right 457 hemifield) and the reward association of the target (reward-associated target vs neutral target). 458 This analysis did not directly test our motivating hypothesis regarding attentional capture to 459 distractor stimuli, but it allowed us to characterize how classification of target location emerged 460 in EEG data and describe the impact of reward association on target processing. Classifiers 461 were trained to identify the location of the lateral target for each of the reward-associated and 462 neutral target conditions separately. As presented in Figure 6a, classification accuracy at the 463 cross-conditional peak (231 – 271 ms) did not reliably differ as a function of the target reward 464 association (p = 0.440). Accuracy subsequently diverges between conditions, but this difference 465 did not survive cluster correction for multiple comparisons.

466 The topographic map presented in Figure 6a illustrates a forward projection of the 467 classification model collapsed across reward-associated and neutral target conditions and 468 averaged over the 231 to 271 ms interval identified by grey shading in the figure. The scalp map 469 shows a lateral pattern, with the model classifying a trial as containing a target in the left 470 hemifield when right posterior cortex had more negative voltage, and vice versa. This suggests 471 that the model is loading on variance that also underlies the N2pc, as has been observed in 472 earlier classification analysis of EEG from visual search (Fahrenfort, Grubert, Olivers, & Eimer, 473 2017). Results from this classification analysis appear to reach a ceiling, with reward 474 association not causing the EEG signal to carry additional information about the target location.





477 478 Figure 8 - Target-elicited ERPs in car / person target conditions. In the stylized stimulus examples here and in Figure 7 the reward-associated category is 'cars', but this was counter-balanced across participants in the actual 479 experiment. Note that here and in Figure 7 negative voltage is plotted upward by convention and the ERPs reflect 480 mean signal observed at the lateral electrode clusters identified by large marker in the topographic maps in Figure 7. 481 a.) The posterior lateral ERPs elicited by a scene containing a peripheral neutral target. The N2pc is apparent as the 482 difference between contralateral and ipsilateral waveforms beginning around 200 ms post-stimulus. b.) The ERPs 483 elicited by a scene containing a peripheral reward-associated target. c.) The ERPs elicited by a scene containing a 484 peripheral neutral target when a task-irrelevant example of the reward-associated category is present in the centre of 485 the scene. d.) The ERPs elicited by a scene containing a peripheral reward-associated target when a task-irreelvant 486 example of the neutral target category is present in the centre of the scene. e.) Contralateral-minus-ipsilateral 487 difference waves for the ERPs illustrated in panels A and B. The N2pc is reflected in negative deflection of the 488 difference wave and does not reliably differ as a function of target reward association, f.) Difference waves for the 489 ERPs illustrated in panels C and D. As in panel E, the N2pc does not reliably differ as a function of target reward 490 association. 491

494 The second classification analysis is analogous to that described above but focussed on 495 classification of distractor location and limited to trials where participants searched for plants. 496 This analysis tests the motivating idea for the study, namely that reward association might 497 impact neural responses to distractor stimuli indexing attentional selection. As illustrated in 498 Figure 6b, distractor location classification emerged quickly and showed a marked difference as 499 a function of whether the distractor category had been associated with reward in prior 500 experience. Across an interval centred on the cross-conditional classification peak (181 - 221 501 ms), accuracy was significantly greater when the distractor was taken from the reward-502 associated stimulus category rather than the neutral stimulus category (p = 0.028). The EEG 503 signal thus carried more information about the location of the reward-associated distractor than 504 it did about the location of the neutral distractor.

505 At first blush, this pattern is consistent with the idea that attention is captured by 506 examples of the reward-associated distractor category. This mis-deployment of attention could 507 cause the EEG signal to carry more information about the distractor location, and this could be 508 associated with the degradation of overt response to the target. However, consideration of the 509 topographic maps illustrated in Figures 6a and 6b identifies an inconsistency in this account. 510 The forward projection of the classification model for distractor localization shows that the model 511 classified a trial as containing a distractor in the left visual hemifield when signal over right 512 posterior cortex had voltage more positive than that over left posterior cortex, and vice versa. 513 This contrasts with results from classification of target location, where contralateral negativity, 514 not positivity, contributed to the model decision.

515 To probe this disparity, we conducted an additional analysis to classify whether a scene 516 contained a central distractor. This was again based on data collected while participants 517 searched for plants, but rather than classifying distractor location, the model labelled trials as 518 either containing a distractor in the centre of the photograph, or not. As illustrated in Figure 6c,



520

521 522 523 524 525 526 527 Figure 9 – Distractor-elicited ERPs in plant target conditions. a.) The posterior lateral ERPs elicited by a scene containing a peripheral neutral distractor. b.) The ERPs elicited by a scene containing a peripheral reward-associated distractor. c.) The ERPs elicited by a scene containing a peripheral neutral distractor when a reward-associated distractor is present in the centre of the scene. d.) The ERPs elicited by a scene containing a peripheral rewardassociated distractor when a neutral distractor is present in the centre of the scene. e.) Contralateral-minus-ipsilateral difference waves for the ERPs illustrated in panels A and B. The N2pc is reflected in negative deflection of the 528 529 difference wave and the P_D is reflected in positive deflection. f.) Difference waves for the ERPs illustrated in panels C and D. Topographic maps reflect the conditional difference in voltage observed in the intervals identified by grey 530 shading in the difference waves.

531

532

the machine learning algorithm was able to perform this task well, with cross-conditional peak decoding accuracy emerging at 176 ms. Topographic projection of the forward model suggests that the classification decision depended on emergence of brain activity in both visual cortex and frontocentral cortex. Classification accuracy was better when the central distractor had been previously associated with neutral outcome than with reward outcome (156 – 196 ms interval; p = 0.025).

540 There is the possibility that distractor location classification and distractor presence 541 classification are related to one another. That is, the contralateral positivity that emerges in 542 classification of distractor location might reflect activation of a mechanism in visual cortex that 543 also emerges when a distractor is present in the center of the scene. Consistent with this, 544 classification of distractor presence appears to depend on bilateral posterior positivity (Figure 545 6c). However, reward association has a positive impact on distractor location classification, but 546 a negative effect on distractor presence classification, and this pattern is hard to explain if 547 classification in both instances is associated with the same EEG variance.

548 An alternative account for this pattern – better classification of the location of a reward-549 associated distractor, but poorer identification of the presence of a reward-associated distractor 550 - is that the reward-associated distractor is suppressed in the post-stimulus interval. Under this 551 premise, classification of the reward-associated distractor is more accurate because this 552 stimulus triggers a response in contralateral visual cortex – a Pd – that acts to inhibit encoding 553 of this stimulus. The machine learning algorithm uses this index of visual suppression to infer 554 distractor location, but separate classification of stimulus presence is poor because this 555 suppression leads to a degraded encoding of the stimulus and its associated category.

556 We conducted two additional analyses of classification accuracy to further test this 557 interpretation. In the first, we examined the latency of peak decoding accuracy for distractor 558 location and distractor presence separately for reward-associated and neutral distractors. Peak 559 classification of the presence of a reward-associated distractor emerged at roughly the same

560 latency as peak classification of the location of a reward-associated distractor (209 ms vs 195 ms; -14 ms difference). However, peak classification of the presence of a neutral distractor 561 562 preceded peak classification of its location (176 ms vs 264 ms; 89 ms difference). Resampling 563 statistics based on these difference scores suggested that the difference in latencies observed 564 for the neutral distractor reliably differed from those observed for the reward-associated 565 distractor (p = 0.038), with follow-up contrasts failing to identify a latency difference between 566 location classification and presence classification for reward-associated distractors (p = 0.276) 567 but identifying a marginal trend for neutral distractors (p = 0.061). This suggests that the EEG 568 signal may carry information about the presence of the neutral distractor that precedes 569 information about its location, consistent with classic theoretical perspectives proposing that 570 diagnostic feature information is extracted from visual input before being localized (Treisman & 571 Gelade, 1980; Wolfe, Cave, & Franzel, 1989). This does not occur for the reward-associated 572 distractor, in line with the idea that brain activity underlying location classification leads to a 573 degraded representation of the distractor, and therefore poorer classification of object presence. 574 In the second analysis, we tested the relationship between the effect of reward on 575 distractor location classification and presence classification. If the EEG variance that supports location classification causes poor classification of distractor presence, there should be a 576 577 negative relationship between these effects across the experimental sample. To this end, we 578 extracted average location and presence classification accuracy across a 100 – 300 ms latency 579 range for each of the high-reward and low-reward conditions for each of the 34 participants. 580 This interval includes latencies where mechanisms of target processing and distractor 581 suppression are known the emerge in the EEG and ERP (Luck & Hillyard, 1994; Hickey, Di 582 Lollo, & McDonald, 2009; Weaver, van Zoest, & Hickey, 2017). As illustrated in Figure 7, as 583 classification accuracy for high-reward distractors increased across individuals (relative to low-584 reward distractors), presence classification decreased (relative to low-reward distractors).

585

586 ERPs

587 Classification of target location appears to rely on emergence of contralateral negativity 588 in posterior cortex, suggesting it is driven by variance underlying the N2pc, whereas 589 classification of distractor location appears to rely on emergence of contralateral positivity, 590 suggesting it is driven by the Pd. However, results from classification leave some ambiguity 591 regarding the contribution of these posterior effects to classification. It is particularly unclear if 592 the difference in distractor location classification as a function of reward association is driven by 593 the Pd or other sources of variance in the EEG signal.

594 To address this ambiguity, and to generally unpack the classification results, we 595 extracted ERPs from the experimental data and isolated the N2pc and Pd components. Figure 8 596 illustrates ERP results when the target of search was a car or person and the target appeared at 597 a lateral location. These analyses do not test our core motivating hypothesis regarding the 598 capture of attention to reward-associated distractors, but, as with classification of target location, 599 allows us to additionally characterize how reward association impacted target processing and to 600 identify the relationship between ERP results and classification. Lateral waveforms are 601 presented in Figures 8A through 8D, and contralateral-minus-ipsilateral difference waves are 602 presented in Figures 8E and 8F. When the lateral target was presented without a central 603 distractor, it elicited a robust N2pc that did not reliably vary as a function of the manipulation of 604 reward outcome (Figure 8E). A smaller N2pc was elicited by the lateral target when a central 605 distractor was present in the scene, reflecting the distracting effect of a prominent foreground 606 non-target at fixation, but, again, the N2pc did not show a reliable effect of target reward 607 association (Figure 8F). The N2pc results are thus very similar to the pattern observed in 608 classification accuracy (Figure 4A). In line with this, the N2pc elicited by a lateral target reliably 609 correlates with target location classification across participants (mean 200 - 300 ms, r = 0.273, 610 p = 0.028, permutation test with 100k iterations).

611 To statistically assess the pattern of results in target-elicited N2pc, we conducted a 612 RANOVA based on mean ERP amplitude from 220 – 280 ms, a latency interval where the N2pc 613 is known to be maximal (Luck & Hillyard, 1994) and where it emerged prominently in the current 614 data. The RANOVA had factors for electrode laterality (ipsilateral vs contralateral), target reward 615 association (reward vs neutral), and central distractor presence (present vs absent), and 616 identified main effects of electrode laterality (F(1,33) = 73.74, p < 0.001, n_{2p}^2 = 0.691), reflecting 617 consistent emergence of N2pc across conditions, and central distractor presence (F(1,33) = 618 5.15, p = 0.030, η_{2p}^{2} = 0.135), reflecting a positive shift in the bilateral ERP when the distractor 619 was present. Electrode laterality interacted with distractor presence (F(1,33) = 33.49, p < 0.001, 620 η^2_p = 0.504), reflecting the increase in N2pc amplitude in the distractor absent condition, but no 621 other interactions emerged (reward X distractor presence: F(1,33) = 1.34, p = 0.255; all other Fs 622 < 1).

623 Figure 9 illustrates ERP results when search was for plants and a reward-associated or 624 neutral distractor appeared at a lateral location. Results from these critical experimental 625 conditions directly address motivating hypotheses regarding the capture of attention to reward-626 related distractors. Lateral waveforms are presented in Figures 9A through 9D, and 627 contralateral-minus-ipsilateral difference waves are presented in Figures 9E and 9F. When the 628 lateral distractor appeared without a central distractor it elicited a lateral response that had both 629 positive and negative components (Figure 9E). This may reflect the imbalance in sensory 630 energy in these scenes; the lateral distractor is a prominent foreground object in these 631 photographs without a corresponding object in the contralateral field, and this imbalance in 632 sensory stimulation may have elicited contralateral activity in visual cortex linked to sensory and 633 perceptual processing unrelated to the deployment of attention. The important observation is 634 that the lateral ERP consistently has more positive polarity in the latency of N2pc and Pd when 635 the eliciting distractor has been associated with reward. This is the case both when the scene 636 contains a central distractor (Figure 9F) and when it does not (Figure 9E). This relative positivity

637 in the lateral ERP suggests that selection of the reward-associated distractor was degraded,638 relative to selection of the neutral distractor.

639 We statistically assessed this pattern in two latency intervals: 100 – 160 ms, which 640 corresponds to the interval when early Pd emerges (Weaver, van Zoest, & Hickey, 2017; 641 Sawaki & Luck, 2010), and 220 – 280 ms, when the N2pc and Pd are maximal (Luck & Hillyard, 642 1994; Hickey, Di Lollo, & McDonald, 2009). In both cases, RANOVA analysis had factors for 643 electrode laterality (ipsilateral vs contralateral), distractor reward association (reward vs neutral), 644 and central distractor presence (present vs absent). Results from the early latency period 645 identify a single main effect of distractor presence (F(1,33) = 11.65, p = 0.002, η_{p}^{2} = 0.261) 646 alongside an interaction of electrode laterality and distractor reward association (F(1,33) = 4.32, 647 p = 0.046, η^2_p = 0.116), reflecting the positive shift in the lateral waveform elicited by a reward-648 associated distractor, and an interaction of electrode laterality and distractor presence (F(1,33) 649 = 15.48, p < 0.001, η^2_p = 0.319), reflecting the negative shift in the lateral waveform when the 650 central distractor was present (electrode laterality: F(1,33) = 3.80, p = 0.060; reward: F(1,33) =651 1.84, p = 0.184; all other Fs < 1). Much the same pattern emerged in analysis of the later time 652 window, with a main effect of electrode location (F(1,33) = 4.13, p = 0.050, η^2_p = 0.111), an interaction of electrode location and reward (F(1,33) = 6.23, p = 0.018, η_p^2 = 0.159), an 653 654 interaction of electrode location and distractor presence (F(1,33) = 13.43, p < 0.001, η_{2p}^{2} = 655 0.289), but no other effects (reward: F(1,33) = 2.13, p = 0.154; distractor presence: F(1,33) =656 2.33, p = 0.134; all other Fs < 1). To relate these ERP effects to classification, we extracted the 657 contralateral signal in the distractor-elicited ERP, collapsing across reward conditions but 658 focussing on the 100 – 300 ms interval where these effects emerged. Increase in positivity in 659 the ERP in this interval reliably predicted an increase in distractor location classification 660 accuracy across individuals (r = 0.304, p = 0.036, permutation test with 100k iterations). This 661 supports the notion that distractor location classification loads on the Pd, as suggested above.

One possibility is that the accentuated suppression of high-reward distractors we identify in Figure 9 is a reaction to the deployment of attention to the reward-associated distractor. However, this hypothesis – deployment of attention to the distractor followed by distractor suppression – should express as an initial contralateral negativity followed by a contralateral positivity (eg. Sawaki & Luck, 2013; Hickey, McDonald, & Theeuwes, 2006). Instead, lateral positivity emerges very quickly, approximately 40 ms after afferent activity reaches visual cortex, leaving little opportunity for preceding cognitive operations.

669 Another possibility is that the reward-associated distractor may initially draw attention 670 within each block, but that participants learn to rapidly suppress this object as they gain 671 experience, either strategically or through the influence of implicit statistical learning (Ferrante, 672 Patacca, Di Caro, Della Libera, Santandrea, & Chelazzi, 2018; Ferrante, Zhigalov, Hickey, & 673 Jensen, 2023). This predicts that the difference in lateral response to neutral and reward-674 associated distractors should change over the 54 trials in a block. In the extreme case, the 675 reward-associated distractor could initially elicit a contralateral negativity - indicative of 676 attentional capture – but later a contralateral positivity, reflecting the establishment of inhibitory 677 control.

678 To test this, we used linear modelling and Bayesian model comparison to assess the 679 impact of trial position within an experimental block. We split observations from each block of 680 trials into two sets, one describing observations from the first half of a block and the other 681 describing those from the second half of a block. If the difference in lateral ERP elicited by 682 reward-associated versus neutral distractors changes over the course of an experimental block, 683 this should emerge as a 3-way interaction of electrode location (ipsilateral vs contralateral), 684 distractor reward association (reward vs neutral), and block position (first half of block vs second 685 half of block). To measure the impact of this 3-way interaction, we repeatedly built mixed linear 686 models for each of the two latency intervals of interest. An initial full model included a random 687 intercept for each experimental participant and fixed factors for reward, distractor presence,

688 electrode location, block position, and all possible interactions between these factors. A 689 restricted model included all these factors, except for the interaction of electrode location, 690 reward, and block position. Bayesian statistics were used to compare the full and restricted 691 models, generating Bayes factor values for each of 1000 iterated model instances that were 692 subsequently mean averaged. Results from analysis of the early latency interval (100 - 160 ms) 693 revealed moderate to strong evidence in favour of the null hypothesis of statistical equivalence 694 of the full and restricted models (average BF = 0.127), and analysis of the late interval 695 generated similar results (220 - 280 ms; average BF = 0.140). The difference in lateral response 696 to reward-associated and neutral distractors - statistically expressed in the interaction of 697 electrode position and reward - therefore appears insensitive to the position of a trial within an 698 experimental block. This suggests that the stimulus-driven suppression indexed in Pd emerges 699 quickly and does not require extended experience of the distractor category.

700

701 Discussion

702 We tested the idea that examples of a reward-associated object category capture attention 703 during search through photographs of real-world scenes. Participants searched for examples of 704 a cued target category - cars, trees, or people - while we recorded electrical brain activity. 705 Importantly, the scenes contained examples of the non-target object categories as task 706 irrelevant distractors. One of the 3 object categories was associated with financial reward, and 707 our interest lay in conditions where search was for a neutral target, but the scene happened to 708 contain an example of the reward-associated category as a task-irrelevant distractor. 709 Behavioural analysis show that participants were slower and less accurate to respond to the 710 target in this circumstance, compared to when the scene contained a neutral distractor. This 711 behavioural pattern has two possible explanations: attention may be captured to the reward-712 associated distractor, or the reward-associated distractor may create filtering costs and the

need for cognitive control. Results from EEG unambiguously show that the behavioural effect is

not a reflection of attentional capture. Instead, it appears that the reward-associated distractor is
suppressed almost immediately after the scene appears. This suppression is indexed in a shift
in distractor-evoked brain activity toward contralateral positivity, indicative of emergence of the
Pd component of the visual ERP (Hickey, Di Lollo, & McDonald, 2009; van Zoest, Huber,
Weaver, & Hickey, 2021), and in degraded accuracy of machine learning classification of
distractor presence.

720 These results contrast with those from existing EEG and MEG studies employing 721 synthetic visual search arrays, where reward-associated distractors appear to robustly capture 722 attention. For example, Hickey, Chelazzi, and Theeuwes (2010) had participants search for a 723 uniquely shaped target in an array of distractors, one of which had unique color. When selection 724 of the target resulted in high magnitude reward, and the target and salient distractor colors 725 subsequently swapped between trials, the distractor captured attention and elicited a robust 726 N2pc. Similarly, Qi et al., (2013) employed a training paradigm to associate reward to a color. 727 When the task changed, and color was rendered task irrelevant, distractors characterized by the 728 reward-associated color continued to capture attention and elicit an N2pc. In these and other 729 studies, reward appears to impact the representation of the task-irrelevant target-characterizing 730 feature such that stimuli with this feature capture attention.

731 Why do reward-associated synthetic distractors capture attention, but reward-associated 732 naturalistic distractors do not? One important observation is that the locus of learning differs 733 across these two contexts. In the studies of synthetic visual search described above, learning 734 presumably involves relatively early visual cortex where low-level features are represented. In 735 studies of naturalistic vision, by contrast, reward is associated to a visually heterogenous 736 category of real-world objects where category membership is not predicted by the presence of 737 specific low-level visual features, and learning impacts encoding in ventral visual cortex, where 738 mid-level features and visual semantics are represented (eq. Hickey & Peelen, 2015, 2017). 739 The association of reward to low-level features may lead to guicker and stronger effects on

740 visual resolution. A second, related observation is that the simplified context of synthetic visual 741 search may provide a better opportunity for incentive salience to cause attentional capture. 742 Relative to naturalistic environments, synthetic search arrays contain only a small set of objects 743 that are characterized by a limited set of non-overlapping visual features. Reward-associated 744 objects may become particularly prominent in this impoverished setting in a way that does not 745 occur in the richer and more complicated context of real-world scenes. Finally, there is the 746 possibility that these differences of perceptual complexity may impact how attentional 747 mechanisms are recruited during search. It may be that the rapid suppression we observe here 748 is only strategically recruited when the visual field contains perceptually complex information 749 with strong competition for limited resources. In line with this, results have shown that reward-750 associated synthetic distractors are also suppressed in early visual cortex, but only when 751 perceptual competition is high (Gong, Jia, & Li, 2017).

752 The idea that reward-associated naturalistic distractors draw attention, but do not 753 necessarily capture it, is broadly in line with a deep literature in visual cognition centred on the 754 idea of signal suppression (eg. Folk & Remington, 1998; Leber & Egeth, 2006; Gaspar & 755 McDonald, 2014; Sawaki & Luck, 2010; see Gaspelin & Luck, 2018, for review). The key 756 proposal here is that salient stimuli may elicit representation in 'salience maps', but that signal at 757 this stage of visual processing can be suppressed so that it does not impact the 'priority maps' 758 that ultimately determine how attention is deployed. For example, in Sawaki and Luck (2010) 759 participants searched through arrays of letters for a target defined by combination of size and 760 character. One of the non-target letters was rendered salient by unique color and results 761 showed that these distractors elicited a prominent Pd component in the ERP, reflecting 762 suppression. As in the current study, this Pd emerged from very soon after stimulus onset, 763 leaving little opportunity for preceding attentional operations. The authors suggested that the 764 salient distractor elicited a salience signal that drew attention to its location. However, because 765 participants knew this salience signal would only identify task-irrelevant stimuli, they strategically

766 inhibited this signal so that the underlying stimulus did not gain selective control (cf. Bacon & 767 Egeth, 1994; see also Sawaki, Geng, & Luck, 2012; Stillwell, Egeth, & Gaspelin, 2022; Drisdelle 768 & Eimer, 2021). The current results suggest that this signal suppression hypothesis can be 769 broadened to describe visual processing of stimuli rendered attention-drawing through reward 770 association. In critical conditions, participants knew that stimuli characterized by incentive 771 salience were task irrelevant. They appear to have been able to establish control through 772 suppression, stopping the deployment of spatial attention and limiting the encoding of 773 information about the task-irrelevant object.

774 If the distractor is suppressed, why is its presence associated with a behavioural cost? 775 Some studies of synthetic visual search find that emergence of distractor suppression is 776 associated with an elimination of behavioural distractor costs (eq. Sawaki & Luck, 2010; 777 Gaspelin & Luck, 2018). However, it is more common to find that distractor suppression reduces 778 distractor costs but does not eliminate them (eq. Burra & Kerzel, 2014; Gaspar & McDonald, 779 2014; Jannati, Gaspar, & McDonald, 2013; Kiss, Grubert, Petersen, & Eimer, 2012). One 780 account for this pattern is that the stimulus-triggered distractor suppression is inefficient. Models 781 of visual attention suggest that the primary purpose of attentional suppression is to shelter 782 neural representations of attended stimuli, limiting interference during the transformation of 783 target information to decisions and behaviour (eq. Desimone & Duncan, 1998; Luck et al., 1997; 784 Tsotsos et al., 1995). If stimulus-triggered suppression is delayed or inefficient, distractors may 785 still interfere with ongoing cognition, though to a lesser degree. A complementary possibility is 786 that residual distractor costs may not reflect interference at all, but rather the cognitive load of 787 stimulus-triggered suppression. By this, effective implementation of stimulus-triggered distractor 788 suppression, which must occur quickly following stimulus onset, may take time and resources, 789 delaying or diminishing the deployment of attention to the target and in this way impacting the 790 speed and accuracy of response (Treisman, Kahneman, & Burkell, 1983; Folk & Remington, 791 1998).

792 To date, the body of neuroscientific literature examining incentive salience in naturalistic 793 visual search has largely relied on fMRI, showing that information about reward-associated 794 distractor categories is degraded in ventral visual cortex (Hickey & Peelen, 2015, 2017; 795 Barbaro, Peelen, & Hickey, 2017; Seidl, Peelen, & Kastner, 2012). This has been counter-796 intuitively interpreted as evidence of the capture of attention to these stimuli. The logic here is 797 that capture will be quick and followed by longer-lived suppression of the distractor to allow 798 search for the target to continue. The notoriously poor temporal resolution of fMRI means that 799 any accentuation of distractor information due to capture is subsumed by the subsequent 800 suppression, and thus that the suppression can be interpreted as a proxy index of capture. The 801 current results challenge this account by showing that the reward-associate distractor is 802 suppressed very quickly in the post-stimulus interval, leaving little opportunity for prior selection. 803 It is important to point out that, even within this new interpretation, evidence of suppression -804 here in the EEG signal, there in the fMRI signal - remains a valid index of the existence and 805 strength of incentive salience. Naturalistic visual objects imbued with incentive salience are 806 attention drawing and need to be strategically suppressed if they are not to be selected. 807 Incentive salience is thought to be of key importance to human addictive behaviour 808 (Robinson & Berridge, 2008). Direct drug stimulation of the midbrain dopamine system is 809 thought to lead to the attribution of incentive salience to drug-related objects and environments. 810 When these objects and environments are encountered in the future, they become difficult to 811 ignore and, once noticed, induce craving and drug-seeking behaviour. In line with this, many 812 studies have reported that task-irrelevant, drug-related stimuli interfere with task-relevant 813 behaviour (see Field & Cox, 2008, for review). However, metanalysis suggests that the 814 relationship between drug craving and attentional bias is not strong (Field, Munafò, & Franken, 815 2009) and the clinical relevance of attentional bias in addictive behaviour is the subject of 816 continuing debate (eq. Christiansen, Schoenmakers, & Field, 2015). This may reflect the 817 mediating influence of strategic attentional control on drug-induced attentional bias. If attentional

818 bias to drug-related stimuli can be reduced through strategic attentional control, drug-related 819 stimuli may be suppressed rather than selected (Garavan & Hester, 2007). This idea is 820 consistent with the broad notion that addictive behaviour is closely linked to reduced activity in 821 inhibitory prefrontal brain regions (Goldstein & Volkow, 2011). It is also consistent with recent 822 results from studies of attentional bias in restrained eating. Though results in the eating 823 literature vary, some studies show that task-irrelevant images of high-caloric foods interfere with 824 strategic behaviour more strongly in un-restrained eaters than in restrained eaters, suggesting 825 that restrained eaters strategically suppress processing of the food stimuli (Forestell, Lau, 826 Gyuvorski, Dickter, & Haque, 2012; Veenstra, de Jong, Koster, & Roefs, 2010; Werthamm, 827 Jansen, & Roefs, 2016; though see Meule, Vögele, & Kübler, 2012; Neimeijer, de Jong, & 828 Roefs, 2012; see Werthamm, Jansen, & Roefs, 2015, for review). A core puzzle in our 829 understanding of addiction and eating disorders is that the same experiences and context can 830 lead to dire disorder in one individual, but leave another unscathed, and there is clear 831 opportunity for research on the strategic attentional control of incentive salience in mediating 832 these outcomes.

To conclude, we demonstrate that prior reward association can cause examples of a category or real-world objects to become salient and attention-drawing. However, these objects do not necessarily capture attention. Participants can establish strategic attentional control over these stimuli, suppressing their representation without the preceding allocation of attention to their location. This neural mechanism for control over incentive salience appears to support adaptive, strategic information-gathering in the natural environment.

- 839
- 840
- 841
- 842
- 843

- 844 References
- Anderson, B. A. (2016). The attention habit: How reward learning shapes attentional
 selection. *Annals of the new York Academy of Sciences*, *1369*(1), 24-39.
- Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional
- capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367-10371.
- 849 Barbaro, L., Peelen, M. V., & Hickey, C. (2017). Valence, not utility, underlies reward-driven
- prioritization in human vision. *Journal of Neuroscience*, *37*(43), 10438-10450.
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation
 and blind deconvolution. *Neural computation*, *7*(6), 1129-1159.
- 853 Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic
- impact, reward learning, or incentive salience?. *Brain research reviews*, *28*(3), 309-369.
- Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.
- 856 Christiansen, P., Schoenmakers, T. M., & Field, M. (2015). Less than meets the eye:
- 857 Reappraising the clinical relevance of attentional bias in addiction. *Addictive*858 *behaviors*, 44, 43-50.
- Della Libera, C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and
 losses. *Psychological science*, *20*(6), 778-784.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial
 EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.
- Besimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.
- 866 Donohue, S. E., Hopf, J. M., Bartsch, M. V., Schoenfeld, M. A., Heinze, H. J., & Woldorff, M. G.
- 867 (2016). The rapid capture of attention by rewarded objects. *Journal of cognitive*
- 868 *neuroscience*, *28*(4), 529-541.

- Fahrenfort, J. J., Grubert, A., Olivers, C. N., & Eimer, M. (2017). Multivariate EEG analyses
 support high-resolution tracking of feature-based attentional selection. *Scientific reports*, 7(1), 1-15.
- 872 Fahrenfort, J. J., Van Driel, J., Van Gaal, S., & Olivers, C. N. (2018). From ERPs to MVPA using
- 873 the Amsterdam decoding and modeling toolbox (ADAM). *Frontiers in Neuroscience*, *12*,874 368.
- Ferrante, O., Patacca, A., Di Caro, V., Della Libera, C., Santandrea, E., & Chelazzi, L. (2018).
 Altering spatial priority maps via statistical learning of target selection and distractor
 filtering. *Cortex*, *102*, 67-95.
- 878 Ferrante, O., Zhigalov, A., Hickey, C., & Jensen, O. (in press). Statistical Learning of Distractor
- 879 Suppression Down-regulates Pre-Stimulus Neural Excitability in Early Visual Cortex. *The*880 *Journal of Neuroscience*.
- Field, M., & Cox, W. M. (2008). Attentional bias in addictive behaviors: a review of its
- development, causes, and consequences. *Drug and alcohol dependence*, *97*(1-2), 1-20.
- 883 Field, M., Munafò, M. R., & Franken, I. H. (2009). A meta-analytic investigation of the
- 884 relationship between attentional bias and subjective craving in substance
- abuse. *Psychological bulletin*, 135(4), 589.
- Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons:
- 887 evidence for two forms of attentional capture. *Journal of Experimental Psychology:*
- 888 Human perception and performance, 24(3), 847.
- 889 Forestell, C. A., Lau, P., Gyurovski, I. I., Dickter, C. L., & Haque, S. S. (2012). Attentional biases
- to foods: the effects of caloric content and cognitive restraint. *Appetite*, *59*(3), 748-754.
- 891 Garavan, H., & Hester, R. (2007). The role of cognitive control in cocaine
- dependence. *Neuropsychology review*, *17*(3), 337-345.
- 893 Gaspar, J. M., & McDonald, J. J. (2014). Suppression of salient objects prevents distraction in
- visual search. *The Journal of Neuroscience*, *34*(16), 5658-5666.

Gaspelin, N., & Luck, S. J. (2019). Inhibition as a potential resolution to the attentional capture
debate. *Current opinion in psychology*, *29*, 12.

897 Goldstein, R. Z., & Volkow, N. D. (2002). Drug addiction and its underlying neurobiological

- basis: neuroimaging evidence for the involvement of the frontal cortex. *American Journal*of *Psychiatry*, *159*(10), 1642-1652.
- 900 Gong, M., Jia, K., & Li, S. (2017). Perceptual competition promotes suppression of reward
- 901 salience in behavioral selection and neural representation. *Journal of*902 *Neuroscience*, *37*(26), 6242-6252.
- Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., & Bießmann, F.
- 904 (2014). On the interpretation of weight vectors of linear models in multivariate
 905 neuroimaging. *Neuroimage*, *87*, 96-110.
- Hershler, O., & Hochstein, S. (2009). The importance of being expert: Top-down attentional
 control in visual search with photographs. *Attention, Perception, & Psychophysics*, *71*(7),
 1478-1486.
- Hickey, C., & Peelen, M. V. (2015). Neural mechanisms of incentive salience in naturalistic
 human vision. *Neuron*, *85*(3), 512-518.
- 911 Hickey, C., & Peelen, M. V. (2017). Reward selectively modulates the lingering neural
- 912 representation of recently attended objects in natural scenes. *Journal of*
- 913 *Neuroscience*, *37*(31), 7297-7304.
- Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward changes salience in human vision via
 the anterior cingulate. *Journal of Neuroscience*, *30*(33), 11096-11103.
- 916 Hickey, C., Di Lollo, V., & McDonald, J. J. (2009). Electrophysiological indices of target and
- 917 distractor processing in visual search. Journal of cognitive neuroscience, 21(4), 760-775.
- 918 Hickey, C., Kaiser, D., & Peelen, M. V. (2015). Reward guides attention to object categories in
- 919 real-world scenes. *Journal of Experimental Psychology: General*, 144(2), 264.

- Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the
 capture of visual attention. *Journal of cognitive neuroscience*, *18*(4), 604-613.
- Hickey, C., Pollicino, D., Bertazzoli, G., & Barbaro, L. (2019). Ultrafast object detection in
- 923 naturalistic vision relies on ultrafast distractor suppression. *Journal of Cognitive*924 *Neuroscience*, *31*(10), 1563-1572.
- 925 Itthipuripat, S., Vo, V. A., Sprague, T. C., & Serences, J. T. (2019). Value-driven attentional
- 926 capture enhances distractor representations in early visual cortex. *PLoS Biology*, *17*(8),
 927 e3000186.
- Jannati, A., Gaspar, J. M., & McDonald, J. J. (2013). Tracking target and distractor processing
- 929 in fixed-feature visual search: evidence from human electrophysiology. *Journal of*930 *Experimental Psychology: Human Perception and Performance*, 39(6), 1713.
- World. *Trends in cognitive sciences*, *23*(8), 672-685.
- 933 Kiss, M., Grubert, A., Petersen, A., & Eimer, M. (2012). Attentional capture by salient distractors
- 934 during visual search is determined by temporal task demands. *Journal of cognitive* 935 *neuroscience*, 24(3), 749-759.
- 236 Le Pelley, M. E., Pearson, D., Griffiths, O., & Beesley, T. (2015). When goals conflict with
- 937 values: counterproductive attentional and oculomotor capture by reward-related
 938 stimuli. *Journal of Experimental Psychology: General*, *144*(1), 158.
- Leber, A. B., & Egeth, H. E. (2006). It's under control: Top-down search strategies can override
 attentional capture. *Psychonomic bulletin & review*, *13*(1), 132-138.
- 941 Luck, S. J. (2014). An introduction to the event-related potential technique. MIT press.
- 942 Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during
 943 visual search. *Psychophysiology*, *31*(3), 291-308.

- Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between
- 945 monkey neurophysiology and human perception: An ambiguity resolution theory of visual 946 selective attention. *Cognitive psychology*, *33*(1), 64-87.
- 947 Meule, A., Vögele, C., & Kübler, A. (2012). Restrained eating is related to accelerated reaction
 948 to high caloric foods and cardiac autonomic dysregulation. *Appetite*, *58*(2), 638-644.
- 949 Mruczek, R. E., & Sheinberg, D. L. (2005). Distractor familiarity leads to more efficient visual
 950 search for complex stimuli. *Perception & psychophysics*, *67*(6), 1016-1031.
- 951 Neimeijer, R. A., de Jong, P. J., & Roefs, A. (2013). Temporal attention for visual food stimuli in
 952 restrained eaters. *Appetite*, *64*, 5-11.
- 953 Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate
- 954 pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in*
- 955 *neuroinformatics*, 10, 27.
- Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural
 basis. *Trends in cognitive sciences*, *18*(5), 242-250.
- Qi, S., Zeng, Q., Ding, C., & Li, H. (2013). Neural correlates of reward-driven attentional capture
 in visual search. *Brain research*, *1532*, 32-43.
- 960 Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons:
- 961 Electrophysiological evidence for an automatic attend-to-me signal. *Attention*,
 962 *Perception*, & *Psychophysics*, 72(6), 1455-1470.
- 963 Sawaki, R., & Luck, S. J. (2013). Active suppression after involuntary capture of
- attention. *Psychonomic bulletin & review*, *20*(2), 296-301.
- 965 Seidl, K. N., Peelen, M. V., & Kastner, S. (2012). Neural evidence for distracter suppression
- 966 during visual search in real-world scenes. *Journal of Neuroscience*, *32*(34), 11812-
- 967 11819.

- 968 Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems
- 969 of smoothing, threshold dependence and localisation in cluster

970 inference. *Neuroimage*, *44*(1), 83-98.

- 971 Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual
- 972 system. *Nature*, *381*(6582), 520-522.
- 973 Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of
- 974 eye movements and attention in real-world scenes: the role of global features in object
 975 search. *Psychological review*, *113*(4), 766.
- 976 Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive*977 *psychology*, *12*(1), 97-136.
- 978 Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling
 979 visual attention via selective tuning. *Artificial intelligence*, *78*(1-2), 507-545.
- 980 Van Zoest, W., Huber-Huber, C., Weaver, M. D., & Hickey, C. (2021). Strategic distractor

981 suppression improves selective control in human vision. *Journal of*

- 982 *Neuroscience*, *41*(33), 7120-7135.
- 983 Veenstra, E. M., de Jong, P. J., Koster, E. H., & Roefs, A. (2010). Attentional avoidance of high-
- 984 fat food in unsuccessful dieters. Journal of Behavior Therapy and Experimental
 985 Psychiatry, 41(3), 282-288.
- Weaver, M. D., van Zoest, W., & Hickey, C. (2017). A temporal dependency account of
 attentional inhibition in oculomotor control. *NeuroImage*, *147*, 880-894.
- 988 Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search.
- 989 Journal of Experimental Psychology: Human Perception and Performance, 29(1), 121.
- 990 Werthmann, J., Jansen, A., & Roefs, A. (2015). Worry or craving? A selective review of
- 991 evidence for food-related attention biases in obese individuals, eating-disorder patients,
- 992 restrained eaters and healthy samples. Proceedings of the Nutrition Society, 74(2), 99-
- 993 114.

994	Werthmann, J., Jansen, A., & Roefs, A. (2016). Make up your mind about food: A healthy
995	mindset attenuates attention for high-calorie food in restrained eaters. Appetite, 105, 53-
996	59.

- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature
 integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, *15*(3), 419.
- Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes
 involves selective and nonselective pathways. *Trends in cognitive sciences*, *15*(2), 7784.

Tree

Lateral Person

Lateral Person

Lateral Car

Bush















(percentage accuracy)



