

Ultrafast object detection in naturalistic vision
relies on ultrafast distractor suppression

Clayton Hickey^{1,2}, Daniele Pollicino², Giacomo Bertazzoli², & Ludwig Barbaro^{1,2}

1 – School of Psychology, University of Birmingham, UK
2 – Center for Mind / Brain Sciences, University of Trento, IT

in press, *Journal of Cognitive Neuroscience*

Word Count Abstract: 150
Word Count Intro: 922
Word Count Discussion: 1065
Number of References: 33

Correspondence:
Clayton Hickey
School of Psychology
University of Birmingham
B15 2TT, Birmingham, UK

c.m.hickey@bham.ac.uk

Abstract

People are quicker to detect examples of real-world object categories in natural scenes than is predicted by classic attention theories. One explanation for this puzzle suggests that experience renders the visual system sensitive to mid-level features diagnosing target presence. These are detected without the need for spatial attention, much as occurs for targets defined by low-level features like color or orientation. The alternative is that naturalistic search relies on spatial attention but is highly efficient because global scene information can be used to quickly reject non-target objects and locations. Here, we use ERPs to differentiate between these possibilities. Results show that hallmark evidence of ultrafast target detection in frontal brain activity is preceded by an index of spatially-specific distractor suppression in visual cortex. Naturalistic search for heterogenous targets therefore appears to rely on spatial operations that act on neural object representations, as predicted by classic attention theory. Participants appear able to rapidly reject non-target objects and locations in order to constrain naturalistic search and increase search efficiency, possibly reflecting the use of global scene information.

Experiment 1 - Discrimination task



Figure 1 – Examples of scene stimuli employed in Experiment 1.

Introduction

In day-to-day life our visual system is flooded with a rich stream of input that we must parse for objects of interest. This real-world selection problem has been approximated in the laboratory in *visual search* experiments in which participants are asked to detect or report characteristics of targets hidden among distractors. To render the issue tractable, visual search experiments have tended to employ simple, synthetic stimuli in regular arrays with targets defined by unique visual features or limited combinations of features.

This reductionist approach has had many notable successes, but it has also led to misunderstanding (Peelen & Kastner, 2014; Wolfe et al, 2011). Classical attention theories - developed from experiments with synthetic stimuli in arrays - suggest that attention creates object representations by binding low-level visual features to each other and to locations. By this, search for complex naturalistic targets requires the serial application of attention so that objects can be compared to target templates (Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989). Search for a naturalistic target in a synthetic array should therefore be more efficient than search for the same target in a visually-complicated natural scene, simply because scenes will tend to contain more objects. But this is not the case. People are extraordinarily quick to detect the presence of real-world objects in scenes (eg. Potter, 1975; Wolfe, Alvarez, Rosenholtz, Kuzmina, & Sherman, 2011; Rousselet, Fabre-Thorpe, & Thorpe, 2002) and neural evidence of target detection emerges in frontal cortex as early as 150 ms post-stimulus (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001; Fabre-Thorpe, Delorme, Marlot, &

Thorpe, 2006). This ultrafast object detection occurs when attention is effortfully engaged elsewhere (Li, VanRullen, Koch, & Perona, 2002; Oliva & Torralba, 2007 for review) and has been linked to aspects of scene context that are ignored by classical theory (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Torralba, Oliva, Castelhana, & Henderson, 2006).

Two proposals have been offered to solve this puzzle. One is that recognition of target presence might rely on the detection of disjunctive sets of features that characterize the target category (eg. Evans & Treisman, 2005; Treisman, 2006). Simple detection of these features may not require attention, but derivation of details about the eliciting object and its location would still rely on the deployment of attention. This predicts that evidence of target detection may emerge in non-visual brain areas without the necessary precedence of spatially-localized, attention-related activity in retinotopic visual cortex.

The alternative is that categorization and detection of naturalistic targets relies on serial spatial selection, but that global scene information rapidly constrains the search space to reduce the number of objects and locations that need be resolved (eg. Torralba, Oliva, Castelhana, & Henderson, 2006; Wolfe et al., 2011). The prediction here is that spatially-localized attentional mechanisms operate prior to the detection of naturalistic targets, but that these spatial effects emerge very quickly.

Here we use human electrophysiology to differentiate between these proposals. To index spatial processing in naturalistic search we looked to lateral components in the visual event-related potential (ERP) - the N2pc and distractor positivity (Pd). The N2pc

is an index of attentional selection that emerges around 175 ms post-stimulus as a negative difference in ERPs recorded over visual cortex contralateral versus ipsilateral to an attended target (Luck & Hillyard, 1994a, 1994b). The Pd is similarly defined as a lateral difference in visual ERPs but has polarity opposite to the N2pc and reflects suppression of unattended stimuli rather than selection (Hickey, Di Lollo, & McDonald, 2009; Sawaki, Geng, & Luck, 2014; Gaspar & McDonald, 2014). The Pd has more temporal variability than the N2pc, emerging earlier when task confines support rapid distractor suppression (Sawaki & Luck, 2010; Kerzel, Barras, & Grubert, 2018; Weaver, van Zoest & Hickey, 2017) and when overt responses are quick (Gaspar & McDonald, 2014).

Few existing studies have employed lateral ERP components as indices of selective processing in scenes. The N2pc has been observed in a set of studies investigating search through scenes, but the target in this work was superimposed over the scene and was not itself naturalistic (Patai, Daollo, & Nobre, 2012; Daollo, Patai, & Nobre, 2013). To our knowledge, only one study has attempted to elicit an N2pc with naturalistic targets - targets defined by their membership in a heterogeneous real-world category - but failed to identify reliable effects (Das, Guo, Giesbrecht, & Eckstein, 2010).

Experiment 1 therefore had the simple goal of characterizing spatial object processing in the lateral ERP during discrimination of naturalistic targets in scenes. We had participants search for examples of people and vehicles in photos of urban scenes in alternating blocks, asking them to discriminate the orientation of the target. As described in the methods below, each scene contained examples of both object

categories, but we prepared the scenes such that we could extract lateralized ERP activity elicited by individual category examples when they acted as target and when they acted as salient distractor. To foreshadow, a robust, normal-latency N2pc is elicited by targets in Experiment 1, and a small, quick Pd is elicited by distractors. Experiment 2 expands from this to investigate the latency of spatially-selective object processing in ultrafast object detection. Experiment 2 had two goals: to verify that the early Pd identified in Experiment 1 emerges during detection of naturalistic targets, and to determine whether this effect precedes emergence of information about target presence elsewhere in the brain. To this end, Experiment 2 contrasts the onset of early Pd with the latency of a hallmark of ultrafast object detection in frontal cortex.

Experiment 1

Methods

Participants

Sixteen healthy University of Trento students gave informed consent before completing the experiment. All reported normal or corrected-to-normal vision and were paid for their participation. One participant was removed from analysis due to an inability to complete the task without moving the eyes, resulting in the rejection of more than 50% of trials due to oculomotor artifacts in the EEG. All remaining participants (5 men; 19 – 26 years old) were right handed. Sample size for both experiments was chosen based on results from pilot experiments suggesting ample power to detect effects of interest.

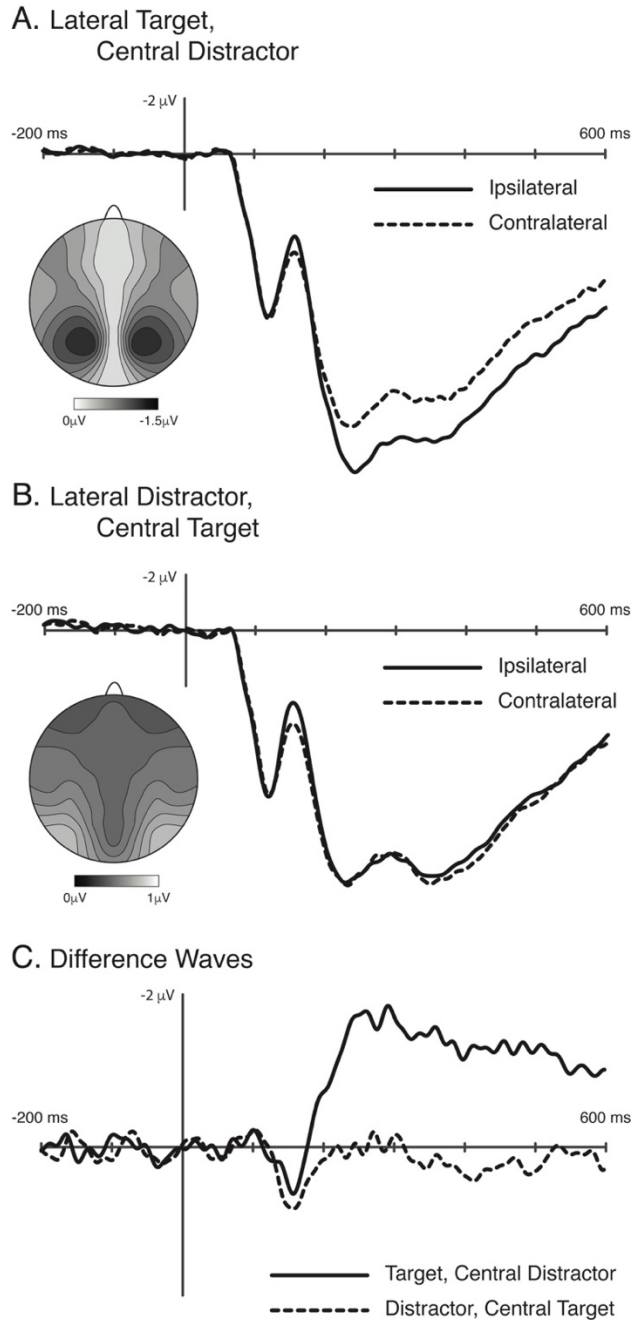


Figure 2 – Waveforms from Experiment 1 as recorded at PO7/8, where both Pd and N2pc were maximal. Note that negative is plotted upward here and in subsequent figures and that stimulus onset occurred at graph origin. Topographic maps reflect mean difference between ipsilateral and contralateral waveforms mirrored across the hemispheres with the midline electrodes set to zero value. A.) ERPs observed ipsilateral and contralateral to a lateral target. Topographic map reflects lateral voltage difference from 200 – 250 ms post-stimulus. B.) ERPs observed ipsilateral and contralateral to a lateral distractor. Topographic map reflects lateral voltage difference from 132 – 182, which reflects a 50 ms. interval centered at the cross-conditional peak of the contralateral positivity at 155 ms. C.) Contralateral-minus-ipsilateral difference waves. The N2pc is reflected in negative waveform polarity and Pd is reflected in positive polarity.

Stimuli and Procedure

Participants were seated in a dim room at approximately 70 cm distance from a 24-inch LED monitor designed for psychophysical research (VPixx Technologies; 100 Hz refresh rate). Each experimental trial began with presentation of a fixation cross for 500 – 1000 ms (randomly selected from rectangular distribution) before one of 480 photographs of complex urban scenes was presented at fixation ($\sim 7^\circ \times \sim 9^\circ$ visual angle; see Figure 1 for examples). Photos sustained until response, when a new trial began. Approximately 80% of photos were taken by authors DP and GB using a standard SLR camera, with the remaining 20% taken from the LabelMe image database (Russell, Torralba, Murphy, & Freeman, 2007). All images were rendered in black and white.

Each photo contained foreground examples of a vehicle and a person with one category example always appearing in the center of the scene and the other appearing in the lateral periphery. Vehicles and people were themselves oriented to the left or the right (ie. side view of vehicle, profile of person) and participants reported the orientation of the target example. Left hand response on a standard keyboard indicated a left-facing target and right hand response indicated a right-facing target. To be clear, response was thus based on the direction that the vehicle or person faced, not on its location in the scene. Photos were selected or prepared such that the category example in the periphery was roughly equidistant from fixation in each image and such that people and vehicles had roughly consistent size across the image set.

Participants completed 40 practice trials before completing 20 blocks of 48 trials. The target category was cued at the beginning of each block and alternated across

blocks (with the order of alternation counterbalanced across participants such that equal numbers started with each of the two target categories). This created a situation where, when search was for people, vehicles were a salient distractor category because they had recently served as targets (and vice versa). Accuracy feedback was provided at the end of each block.

Stimuli presented at fixation or along the vertical meridian above and below fixation are processed equally by both visual cortices and therefore do not create lateralized activity in the visual ERP (Woodman & Luck, 2003; Hickey, McDonald, & Theeuwes, 2006; Hickey, Di Lollo, & McDonald, 2009). As a result, when the target appeared in the periphery and the distractor appeared at fixation, lateralized brain activity could be unambiguously associated with target processing. Equally, when the salient distractor appeared in the periphery and the target appeared at fixation, lateral brain activity could be linked to distractor processing.

Participants saw each photo twice, once when the vehicle was a target and once when the person was a target, and the experiment took approximately 1 hour to complete. All procedures were approved by the University of Trento ethics commission.

EEG Recording and Analysis

EEG was recorded at 1 kHz from 64 Ag/AgCl electrodes mounted in an elastic cap (10/20 montage) using the BrainAmp-DC system and Brain Vision Recorder software (Brain Products). Additional electrodes were placed at the left and right mastoids and 1

cm lateral to the outer canthus of the left eye. All electrodes were referenced during recording to the right mastoid and subsequently re-referenced offline to the algebraic average of the mastoids. Electrode impedances were kept below 10 k Ω and EEG was digitally filtered offline at 0.01 – 45 Hz to remove direct current drift and high-frequency noise (Blackman-windowed FIR-filter, 10k kernel length).

Infomax independent component analysis (ICA; Bell & Sejnowski, 1995) was used to identify variance stemming from ocular artifacts in the EEG. The ICA components representing horizontal and vertical eye movements was used to identify trials in which eye movements were made in the 600 ms interval following stimulus onset. Participants retained for analysis moved their eyes in 6 – 10% of trials; these trials were removed from analysis. Components representing eye artefacts were subsequently removed from the data and ERPs were calculated based on trials where participants made correct responses. ERPs were baselined on the interval beginning 100 ms before stimulus onset and ending 50 ms after.

Statistical analysis

Contrasts of behaviour and ERP components measured over predefined intervals relied on bootstrap analysis of difference scores with 10^5 iterations. Effect size for these contrasts is reported as Cohen's d (Cohen, 1988). Where noted, we additionally assessed lateral brain activity using a randomization method introduced by Sawaki, Geng and Luck (2012). This tests the hypothesis that lateral ERP amplitude in a given interval is larger than expected by chance. Critically, only directional component

amplitude is analyzed, generating a measure of signed area in the difference wave. That is, the N2pc is measured by summing the negative polarity signal in the contralateral-minus-ipsilateral difference across a long, predefined interval (50 – 500 ms in the current experiment). This approach is adopted because it allows the N2pc to be measured and assessed without the need for *a priori* definition of precise temporal analysis windows, even when the N2pc occurs in close temporal proximity to opposite-polarity effects like the Pd (which will cancel out N2pc amplitude when included in mean averages). However, because the signed area approach ignores any finding that is contrary to the hypothesis under consideration – that is, when the negative-polarity N2pc is being tested, it ignores any positive polarity amplitude – it is a biased estimator. To test if a biased estimate of an effect is significant, it must be compared against a distribution of results observed when no effect is present but the bias in estimation remains.

To generate such a null distribution, we iteratively relabeled ipsilateral and contralateral waveforms for each participant, took a measure of mean negative polarity in the difference wave calculated across the resulting datasets, and stored this in an accumulator. We did this for all 2^{15} combinations of ipsilateral and contralateral waveforms in the 15-participant dataset. When the observed effect lay at the extreme edge of this null distribution, it was unlikely to be a product of chance and the null hypothesis could be discarded. The associated *p* value was calculated as the proportion of the null distribution that lay at or beyond this observation. The same approach was

adapted for statistical assessment of Pd, but with the measurement of positive rather than negative polarity.

We adapted this approach to test the difference in Pd observed in two experimental conditions. To create a null distribution for this test, we generated the contralateral-minus-ipsilateral difference wave for all 2^{15} combinations of ipsilateral and contralateral waveforms for each of the two conditions of interest separately and took a measure of positive polarity for each difference wave. We subsequently calculated the difference between these biased estimators for all 2^{30} combinations of results for each of the two conditions. Importantly, at this stage both positive and negative polarity differences were included in the null distribution.

Results

Behavioural Results

When the target appeared at a lateral location and the distractor appeared at fixation, mean reaction time (RT) was 600 ms (62 ms SD) and mean error rate was 4.7% (3.8% SD). When the target was central and the distractor lateral, participants were both reliably faster (558 ms, 59 ms SD; $p < 10^{-5}$, $d = 0.70$) and more accurate (2.5% error, 2.1% SD; $p < 10^{-5}$, $d = 0.72$).

Electrophysiological Results

In the lateral target condition – when the distractor was presented at fixation – signed area analysis revealed a robust target-elicited N2pc beginning at approximately 200 ms.

post-stimulus ($p < 10^{-4}$; Figure 2A). This was preceded by apparent emergence of an early, positive-polarity difference at around 150 ms post-stimulus. This may reflect a sensory imbalance in the scenes: the prominent foreground presence of the lateral target in one field was not balanced by a corresponding foreground stimulus in the opposite field, and this may have created sensory laterality in the visual ERP. However, this positive polarity effect was not statistically significant in signed area analysis ($p = 0.666$).

In the lateral distractor condition – when the target was presented at fixation – the early positivity appears larger in amplitude and is followed by a hint of continued positive signal (Figure 2B). Because the raw ERP is dominated by large positive deviation in both ipsilateral and contralateral waveforms, these positive polarity effects are more clearly visible in the difference waves provided in Figure 2C. Though the absolute lateral positivity was only marginally reliable in signed area analysis ($p = 0.090$), the waveform was reliably more positive in the lateral distractor condition than in the lateral target condition, both when measured using the signed area approach over a large temporal interval (50 – 500 ms; $p = 0.048$) and using a more traditional measure of mean amplitude across a 40 ms interval centered on the 155 ms cross-conditional peak of positive polarity ($p = 0.015$, $d = 0.322$).

Existing research has associated this early positive polarity effect – the early Pd – with suppression of the lateral distractor (Sawaki & Luck, 2010; Kerzel, Barras, & Grubert, 2018; Weaver, van Zoest & Hickey, 2017). Its emergence in Experiment 1 thus constitutes an indicator of rapid distractor suppression during naturalistic search. As

Experiment 2 - Detection task



Figure 3 – Examples of scene stimuli employed in Experiment 2.

detailed in the introduction, Experiment 2 was designed with two primary purposes: a.) to test if the early Pd emerges during naturalistic target detection, and b.) to determine if it shows variance as a function of target presence prior to the emergence of evidence of target presence in frontal cortex.

Experiment 2

Methods

Participants

Sixteen healthy University of Trento students gave informed consent before completing Experiment 2. None had taken part in Experiment 1, all reported normal or corrected-to-normal vision, and all were paid for their participation. Fourteen of the participants (7 men; 18 – 26 years old) were right handed.

Stimuli and Procedure

A new set of 600 photographs was employed in Experiment 2. This set partially overlapped with that employed in Experiment 1 but was supplemented by additional photos taken by authors DP and GB and by additional images from the LabelMe database (Russell, Torralba, Murphy, & Freeman, 2007).

Photos once again contained examples of people and vehicles, but in Experiment 2 participants had to report whether an example of the target category was present in the scene. Half of participants responded with the left hand when the target was present and the right hand when it was not with the reverse response map for the remaining participants.

Scenes contained a target in 66% of trials. In 200 photos, a person appeared to the left or right and a vehicle appeared at fixation, in 200 photos a vehicle appeared to the left or right and a person appeared at fixation, in 100 photos a vehicle appeared to the left or right with no person in the scene, and in 100 photos a person appeared to the left or right with no vehicle in the scene. Unlike in Experiment 1, vehicles and people could face directly toward or away from the camera. In order to increase task demands on selective attention, the scene was presented for 60 ms before being replaced by a

naturalistic mask (created by generating white noise at a wide range of spatial frequencies and superimposing a naturalistic texture; see Hickey & Peelen, 2015, Figure 1, for an illustration).

Participants completed 60 practice trials before completing 20 experimental blocks of 60 trials. Images containing both object categories appeared twice during the experiment, once when participants were instructed to search for vehicles and once when they were instructed to search for people. All other procedures, including all EEG recording and analysis, were as in Experiment 1.

Statistical Analysis

In addition to the statistical approaches detailed for Experiment 1, in Experiment 2 lateral ERP component onset and offset latencies were measured and compared. For comparisons between components, onset and offset latencies were defined as the time at which an effect either first surpassed 80% of maximum or first fell below 80% of maximum. To test the reliability of conditional differences in onset and offset latency, null distributions were iteratively generated by randomly relabeling observations for each of the two effects of interest. That is, in a comparison of the latency of effect A to effect B, the null distribution was created by iteratively relabeling effect A and effect B for subsets of participants, measuring the latency difference between mean averages of these two synthetic sets, and adding this to an accumulator. This was carried out for all 2^{16} combinations of effect A and effect B and the observed effect was compared to this null distribution.

In analysis of Experiment 2 we additionally compare the latency of a conditional difference in Pd components to the latency of a conditional difference in frontal ERPs. Because these effects differed in magnitude, the percent-of-maximum approach to measuring onset was unsuitable (i.e. the much-larger frontal effect was present for some time before reaching a substantive percentage of its maximum value). To estimate the onset of these effects we contrasted the experimental conditions for each 1 ms sample in the ERP to identify the first evidence of statistically reliable difference (uncorrected threshold of $p < 0.01$). The first sample showing this difference that was followed by 4 more samples showing the same difference was defined as the onset of the effect. It is important to note that we employed non-causal filters during digitization of the EEG and thereafter, and that these filters can shift effects forward in time. This is particularly the case when effects are large in magnitude, as is the case for frontal effects in the current data. The statistical estimates of absolute onset generated by this analysis should therefore be understood to be biased toward early estimation, particularly in the case of the large frontal effect. However, because the same filters are applied to all data, differences in onset observed between conditions are minimally impacted. Moreover, in the particular circumstance observed in the current results - where the larger effect follows the smaller effect - this impact of filtering will make estimates of latency differences between conditions conservative.

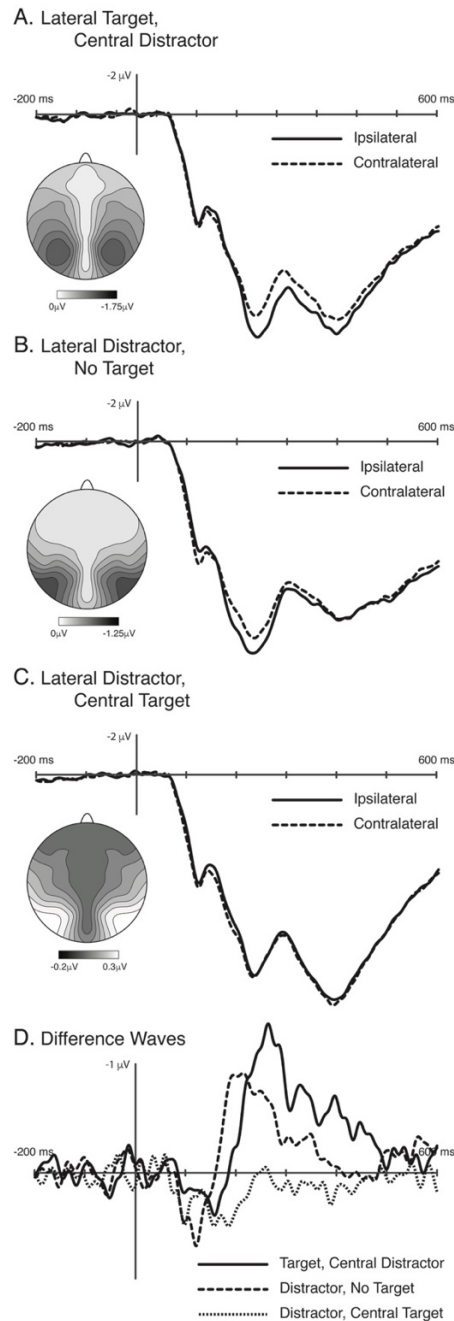


Figure 4 – Waveforms from Experiment 2 as recorded at PO7/8, where both Pd and N2pc were maximal. A.) ERPs observed ipsilateral and contralateral to a lateral target when a distractor appeared at fixation. Topographic map reflects lateral voltage difference from 239 – 289 ms post-stimulus, which reflects a 50 ms. interval centered at the N2pc peak. B.) ERPs observed ipsilateral and contralateral to a lateral distractor when no target appeared in the scene. Topographic map reflects lateral voltage difference from 189 – 239 ms post-stimulus, which reflects a 50 ms. interval centered at the N2pc peak. C.) ERPs observed ipsilateral and contralateral to a lateral distractor when the target appeared at fixation. Topographic map reflects lateral voltage difference from 160 – 210 ms post-stimulus, which reflects a 50 ms. interval centered at the Pd peak. D.) Contralateral-minus-ipsilateral difference waves.

Results

Behavioural Results

When the target appeared at a lateral location and the distractor was central, mean reaction time (RT) was 585 ms (108 ms SD) and miss rate was 21% (16% SD).

Participants were reliably faster (539 ms, 105 ms SD; $p < 10^{-5}$, $d = 0.424$) and more accurate (2.2% misses, 2.4% SD; $p < 10^{-5}$, $d = 1.18$) when the target was central and the distractor was lateral. The slowest responses occurred when the target was absent from the scene (642 ms, 119 ms SD) where a false alarm rate of 7.9% was observed (4.7% SD).

Electrophysiological Results

In the lateral target / central distractor condition, when the distractor was presented at fixation, signed area analysis revealed a target-elicited N2pc that began at approximately 200 ms post-stimulus ($p = 0.001$; Figure 2A). As in Experiment 1, this was preceded by early, positive-polarity activity that potentially reflects an imbalance in the visual properties of the scene, here emerging at around 100 ms, but this did not emerge reliably from signed area analysis ($p = 0.589$). Similar results were observed in the lateral distractor / no target condition, with the distractor rather than the target as the eliciting stimulus. A robust distractor-elicited N2pc emerged at approximately 180 ms ($p = 0.002$; Figure 2B) and was preceded by a positive effect at around 100 ms that did not reach significance ($p = 0.616$). In the lateral distractor / central target condition, when the target was presented at fixation, the lateral ERP shows no hint of a distractor-

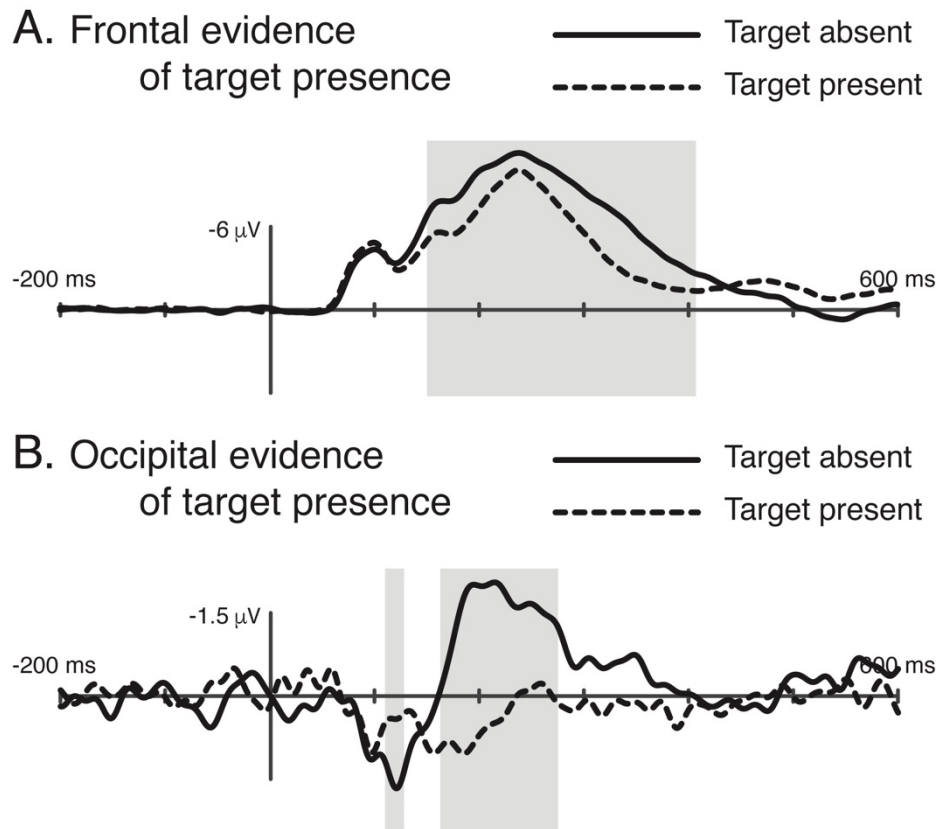


Figure 5 – A.) Fronto-central ERP as a function of whether target was present or absent in eliciting scene. B.) Lateral occipital contralateral-minus-ipsilateral difference wave as a function of whether target was present or absent in eliciting scene. Note that this panel reproduces Figure 4D for the purpose of illustration and comparison. Samples showing statistically significant difference between the waveforms are identified by shaded background ($p < 0.01$). No earlier latency difference was identified when a more liberal criterion was adopted ($p < 0.05$).

elicited N2pc, but the early positivity sustains until later in the ERP and signed area analysis identified a reliable Pd ($p = 0.001$).

Onset analysis showed that the distractor-elicited N2pc in the lateral distractor / no target condition emerged 54 ms earlier than the target-elicited N2pc in the lateral target / central distractor condition ($p < 10^{-4}$). Analysis of offset latency demonstrated that the early Pd sustained in the lateral distractor / central target condition 34 ms longer

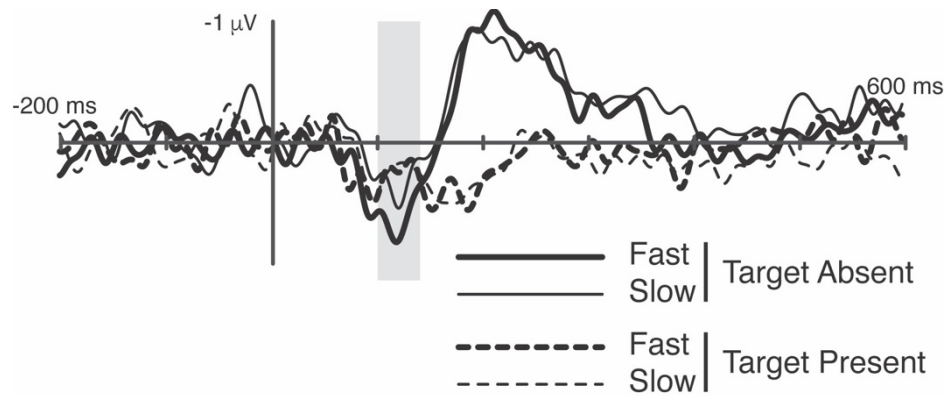


Figure 6 – Contralateral-minus-ipsilateral difference waves as a function of target presence and median split based on trial RT.

than in the lateral target / central distractor condition ($p = 0.005$), which itself sustained 36 ms longer than in the lateral distractor / no target condition ($p = 0.020$).

As illustrated in Figure 5A, the peak of early Pd amplitude in the lateral distractor / no target condition (identified in Figure 5 as the target absent condition) appears of greater magnitude than in the lateral distractor / central target conditions (identified in Figure 5 as the target present condition). To test this, we determined the peak latency of positive lateral effect when these conditions were collapsed (121 ms) and subsequently extracted mean amplitude for each condition over a 40 ms interval centered at this time. The early distractor-elicited early Pd was reliably larger when the target was absent from the scene ($p = 0.033$, $d = 0.590$).

To determine if the early Pd preceded or followed non-spatial evidence of target presence in frontal cortex, we contrasted target-present and target-absent scenes to extract the frontal ERP measure of ultrafast target detection first reported by Thorpe, Fize, and Marlot (2006). This ERP effect expresses as an increase in amplitude when

the eliciting scene does not contain a target and is therefore thought to reflect emergence of a control process involved in stopping a prepotent ‘target present’ response when this would be an error (Thorpe, Fize, & Marlot, 2006). Consistent with this interpretation, it has latency and distribution similar to that of the ‘no-go N2’ (Pfefferbaum, Ford, Weller, & Kopell, 1985; see Folstein & Van Petten, 2008, for review). We measured this effect across a set of frontal midline electrodes (F1, Fz, F2, FC1, FCz, and FC2) where it is known to appear first and where it is maximal (Thorpe, Fize, & Marlot, 2006). As illustrated in Figure 5A, statistical onset analysis determined that target presence emerged in this signal beginning at 150 ms post-stimulus, closely replicating the 152 ms onset identified by Thorpe, Fize, and Marlot (2006). In contrast, target presence impacted the early Pd from 111 ms post-stimulus (Figure 5B).

To probe the functional role of the early Pd, we conducted a median split of conditional results based on trial RTs. As illustrated in Figure 6, the early Pd appears larger when the target was absent from the scene and participants were quick to correctly report its absence. To test the reliability of this pattern, we extracted mean Pd amplitude for each of the 4 waveforms illustrated in Figure 6 using the same 40 ms window centered described above. In the target absent condition the early Pd was larger when participants responded quickly ($p = 0.010$, $d = 0.642$). The interaction of target presence and participant response speed was marginally significant ($p = 0.065$). Note that while amplitude of the early Pd varied as a function of response latency, neither the latency nor amplitude of the subsequent distractor-elicited N2pc showed any difference in this analysis.

General Discussion

What we know of visual search from studies employing synthetic stimuli arrays does not entirely match what we know of visual search from studies employing natural scenes.

One conspicuous inconsistency is the speed of naturalistic search. Classic theories of search – developed from work with synthetic stimuli – suggest search for naturalistic targets in scenes should be slow because scenes contain so many objects. But naturalistic search is blazingly quick.

This disparity has been explained in two ways. One possibility is that detection of naturalistic targets – examples of real-world object categories – relies on the ability to detect any of a set of diagnostic mid-level features. By this, we can quickly infer the presence of a car in a scene because we have detected the combined presence of mid-level features like metallic sheen, a set of wheel-like circles, or a prototypical arrangement of headlights in a grill. This gives us a probabilistic notion that a car is present, with details and information about location filled in only when attention is deployed (Evans & Treisman, 2006).

This account depends on the plausible idea that extended experience of real-world object categories might render our visual system sensitive to mid-level visual features in the same way it is sensitive to low-level features. As a result, detection of overlearned mid-level features may not require attention, just as detection of low-level features – the color red, vertical orientation, movement – does not require attention (Treisman & Gelade, 1980). For the purpose of the present study, the critical prediction of this account is that evidence of ultrafast object detection, like the frontal ERP effect

reported by Thorpe, Fize, & Marlot (1996), will precede evidence of spatial object processing such as distractor suppression.

The alternative is that detection of naturalistic targets requires selective processing, as predicted by classical theory, but that naturalistic search is more efficient than synthetic search. This efficiency could be supported by scene context, which might constrain search to locations likely to contain a target (eg. Torralba, Oliva, Castelhana, & Henderson, 2006; Wolfe et al., 2011). The prediction from this account is that ultrafast object detection should follow spatially-localized object processing, and distractor suppression in particular, but that such localized object processing will emerge very quickly in naturalistic vision.

We used ERPs to differentiate between these accounts. In Experiment 1 we identified an index of early, spatially-localized object processing that emerges during discrimination of naturalistic targets in scenes. This expressed as a rapid positive increase in visual ERP activity contralateral to a salient distractor when the target is absent from the display. This importantly emerged under circumstances where the presence of the target itself could not contribute to the lateral ERP. Earlier research has shown that this early Pd reflects the suppression of distractors (Sawaki & Luck, 2010; Kerzel, Barras, & Grubert, 2018). For example, in concurrent EEG and eye-tracking work, the early distractor-elicited Pd precedes eye movements to targets, is absent when the eyes are captured to distractors, and has magnitude that predicts how close to the center of the target the eyes will land (Weaver, van Zoest, & Hickey, 2017).

The early Pd we observe in Experiment 1, though reliable in the sample, is unarguably small. We believe this is because the target in the eliciting scenes appeared at fixation and was therefore very easy to find. This design feature was adopted to maintain naturalism in the scene and to reduce potential confounds, but it is likely to have rendered the distractor less conspicuous and therefore less demanding of distractor suppression. Some support for this notion is provided by results from Experiment 2, where the early Pd is larger when the target is absent from the scene and attention is more heavily taxed.

The early Pd identified in Experiment 1 constitutes an early index of spatially-localized distractor suppression during search. Experiment 2 builds from this finding to test the involvement of this distractor suppression in the detection of naturalistic targets. A core goal of Experiment 2 was to determine if this kind of object processing was evident in the visual ERP before or after evidence of target presence emerged in frontal brain activity.

Detection of a naturalistic target in Experiment 2 was associated with a robust N2pc, but this component emerged later than the onset of information about target presence in frontal ERPs, and therefore could not play the causative role in target detection that is suggested by classic attention theory. However, frontal evidence of target presence, emerging at 150 ms post-stimulus, was preceded by sensitivity to target presence in the early Pd at 111 ms. The early Pd was larger when the eliciting scene did not contain a target, suggesting an increase in distractor suppression when the target was not easily found. This effect was accentuated when participants quickly

reported the absence of the target, consistent with prior observations of an association between Pd and response latency (Weaver, van Zoest, & Hickey, 2017; Gaspar & McDonald, 2014).

The Pd was briefer when the target was absent from the display, with subsequent emergence of a distractor-elicited N2pc reflecting attentional selection of this stimulus. One possibility here is that participants attentionally selected the distractor in the process of confirming the absence of the target from the scene. However, if this distractor-elicited N2pc were involved in evidence accumulation leading to a target absent response, it might be expected to vary as a function of response latency. But no hint of this emerged in data analysis. The functional role of this distractor-elicited N2pc thus remains an open question: it may be involved in instrumental decision-making, but it does not vary according to expectations derived from this interpretation.

Our results are consistent with the notion that ultrafast object detection relies on spatial operations that act on neural object representations, and this is at odds with conclusions drawn by Battistoni, Kaiser, Hickey, and Peelen (2018) from multivariate analysis of magnetoencephalographic (MEG) data. Participants in Battistoni et al. (2018) completed two tasks. In the first, they detected the presence of a synthetic target rendered salient through manipulation of global display characteristics. In the second, they detected the presence of lateral examples of real-world object categories in photographs of natural scenes. MEG was recorded in both tasks, and a multivariate classifier trained on the synthetic search task was used to determine the timing of selection in the naturalistic search task. Results showed that reliable decoding of target

location emerged only after reliable decoding of target presence, leading the authors to conclude that target detection preceded the action of selective attention. But the effect identified in the current paper – in which target presence changes the action of a spatial mechanism of distractor suppression – does not provide information about target location. The dependent measure employed by Battistoni et al. (2018) was therefore insensitive to the effect identified in the current paper. There is the opportunity for future work employing multivariate classification to identify the earliest emergence of distractor suppression in naturalistic vision.

In summary, we provide evidence that ultrafast object detection involves ultrafast suppression of salient distractors. The quick distractor suppression we identify is consistent with the notion that global scene information can be used to rapidly suppress salient objects that are unlikely to be targets (or, perhaps, locations that are unlikely to hold targets). This mechanism improves efficiency by constraining the search space, giving us the ability to very rapidly detect the presence of targets in natural scenes.

Bibliography

Battistoni, E., Kaiser, D., Hickey, C., & Peelen, M. V. (2018). The time course of spatial attention during naturalistic visual search. *Cortex*.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6), 1129-1159.

Cohen J. (1988). Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic

Das, K. Guo, F. Geisbrecht, B. Eckstein, M. P. (2010). Predicting contextual locations in natural scenes from neural activity [Abstract]. *Journal of Vision*, 10(7):1295.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1), 193-222.

Doallo, S., Patai, E. Z., & Nobre, A. C. (2013). Reward associations magnify memory-based biases on perception. *Journal of cognitive neuroscience*, 25(2), 245-257.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free?. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1476.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of cognitive neuroscience*, 13(2), 171-180.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, 45(1), 152-170.

Gaspar, J. M., & McDonald, J. J. (2014). Suppression of salient objects prevents distraction in visual search. *Journal of neuroscience*, 34(16), 5658-5666.

Hickey, C., & Peelen, M. V. (2015). Neural mechanisms of incentive salience in naturalistic human vision. *Neuron*, 85(3), 512-518.

Hickey, C., Di Lollo, V., & McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *Journal of cognitive neuroscience*, 21(4), 760-775.

Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of cognitive neuroscience*, 18(4), 604-613.

- Kerzel, D., Barras, C., & Grubert, A. (2018). Suppression of salient stimuli inside the focus of attention. *Biological psychology*, 139, 106-114.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, 99(14), 9596-9601.
- Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31(3), 291-308.
- Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1000.
- Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive psychology*, 33(1), 64-87.
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in cognitive sciences*, 11(12), 520-527.
- Patai, E. Z., Doallo, S., & Nobre, A. C. (2012). Long-term memories bias sensitivity and target selection in complex scenes. *Journal of cognitive neuroscience*, 24(12), 2281-2291.
- Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends in cognitive sciences*, 18(5), 242-250.
- Pfefferbaum, A., Ford, J. M., Weller, B. J., & Kopell, B. S. (1985). ERPs to response production and inhibition. *Electroencephalography and clinical neurophysiology*, 60(5), 423-434.
- Potter, M. C. (1975). Meaning in visual search. *Science*, 187(4180), 965-966.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature neuroscience*, 5(7), 629.
- Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1-3), 157-173.
- Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception, & Psychophysics*, 72(6), 1455-1470.

Sawaki, R., Geng, J. J., & Luck, S. J. (2012). A common neural mechanism for preventing and terminating the allocation of attention. *Journal of Neuroscience*, 32(31), 10725-10736.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *nature*, 381(6582), 520.

Torralba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review*, 113(4), 766.

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual cognition*, 14(4-8), 411-443.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97-136.

Vanrullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of cognitive neuroscience*, 13(4), 454-461.

Weaver, M. D., van Zoest, W., & Hickey, C. (2017). A temporal dependency account of attentional inhibition in oculomotor control. *Neuroimage*, 147, 880-894.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, 73(6), 1650.

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), 77-84.

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 29(1), 121.