Nako, R. and Smith, Tim J. and Eimer, Martin (2015) Activation of new attentional templates for real-world objects in visual search. Journal of Cognitive Neuroscience 27 (5), pp. 902-912. ISSN 0898-929X.

Running Head: Activation of new attentional templates

Activation of new attentional templates for real-world objects in visual search

Rebecca Nako, Tim J. Smith & Martin Eimer

Department of Psychological Sciences, Birkbeck College, University of London

Address correspondence to:

Rebecca Nako
Department of Psychological Sciences,
Birkbeck College,
University of London,
UK
beccanako@gmail.com

Abstract

Visual search is controlled by representations of target objects (attentional templates). Such templates are often activated in response to verbal descriptions of search targets, but it is unclear whether search can be guided effectively by such verbal cues. We measured event-related potentials (ERPs) to track the activation of attentional templates for new target objects defined by word cues. On each trial run, a word cue was followed by three search displays that contained the cued target object among three distractors. Targets were detected more slowly in the first display of each trial run, and the N2pc component (an ERP marker of attentional target selection) was attenuated and delayed for the first relative to the two successive presentations of a particular target object, demonstrating limitations in the ability of word cues to activate effective attentional templates. N2pc components to target objects in the first display were strongly affected by differences in object imageability (i.e., the ability of word cues to activate a target-matching visual representation). These differences were no longer present for the second presentation of the same target objects, indicating that a single perceptual encounter is sufficient to activate a precise attentional template. Our results demonstrate the superiority of visual over verbal target specifications in the control of visual search, highlight the fact that verbal descriptions are more effective for some objects than others, and suggest that the attentional templates that guide search for particular real-world target objects are analog visual representations.

Keywords: Attentional selection, N2pc, visual search, object search, spatial attention, attentional control

Introduction

When we look for a particular target object in a crowded visual environment, search is controlled by our knowledge about the visual properties of this particular target. Such representations of task-relevant objects or object features (*attentional templates*) are assumed to reside in visual working memory (e.g., Duncan & Humphreys, 1989; Wolfe & Horowitz, 2004; Olivers et al., 2011). Attentional templates are often described as “images in the mind” (James, 1890), which implies that they are analog visual representations of target objects (e.g., mental images as described by Kosslyn, 1987; see also Kosslyn & Thompson, 2003), rather than abstract propositional representations (e.g., Pylyshyn, 2002). Search templates can be activated prior to the start of visual processing, and facilitate the selection of targets among distractors by guiding attention towards the location of template-matching objects in the visual field (e.g., Desimone & Duncan, 1995; Wolfe, 1994, 2007; Eimer, 2014). Although attentional templates play a central role in models of selective visual attention and visual search, the processes that are involved in the formation of a particular search template have so far rarely been investigated. Most visual search experiments require observers to search for the same target feature or object across many experimental trials, which are typically preceded by practice trials where the visual features of the target object are learned. In such situations, target selection is controlled by a fully established attentional template for a particular target object that remains unchanged throughout the experiment. Visual search in real-world environments is seldom like this. In naturalistic contexts, we rarely look for the same target object repetitively across search episodes, but usually search for one particular target object, and then start search for a different object. Moreover, real-world attentional templates do not always provide an exact match with the visual properties of a particular target object. Search episodes are frequently initiated by verbal instructions (“can you find my bag in the wardrobe?”), which may not constrain the visual features of a target object as precisely as a visual image of the search target (e.g., bags come in different, shapes, colors, or sizes).

If attentional templates are visual representations of target objects, search should be guided more efficiently once a target has been encountered visually than when its identity is specified only by verbal description. Such a difference was indeed observed by Wolfe et al. (2004) in a study where search targets changed across successive trials, and the identity of each target was indicated at the start of each trial by a picture cue or a word cue. Each cue display was followed by a single search display, and the stimulus onset asynchrony (SOA) between these two displays was varied. Targets were detected faster as SOAs became longer, demonstrating that the activation of an attentional template for a new target object does not happen instantaneously, but is a time-consuming process (see also Dombrowe, Donk, & Olivers, 2011). Importantly, Wolfe et al. (2004) found that the speed with which a new attentional template could be implemented differed markedly between picture and word cues (see also Wolfe, Butcher, Lee, & Hyle, 2003, Vickery, King, & Jiang, 2005, and Schmidt & Zelinsky, 2009, for similar observations). When the picture cue was an exact image of the search target, attentional templates were set up rapidly, within about 200 ms. When target identity was signaled by a word cue (e.g., “black vertical” or “rabbit”), the activation of an attentional template was slower, and target selection remained less efficient than with picture cues even with long cue-target SOAs (800 ms). Similar performance differences between picture and word cues were observed regardless of whether observers searched for targets defined by conjunctions of simple features (e.g., black vertical bars) or for images of real-world objects (e.g., rabbits).

While such behavioral findings demonstrate that attentional templates guide visual search more effectively when target identity is specified by images rather than words, they do not provide direct insights into which stages of attentional processing are affected by this difference between visual and verbal target definitions. Does the initial spatial selection of target objects operate more rapidly when their identity is signaled by picture cues as compared to word cues, or are the performance advantages observed with picture cues primarily generated at later target identification stages? In the current study, we combined behavioural and electrophysiological measures to track the speed and efficiency of selecting a target object defined by a word cue in real time, and to contrast these selection processes with processes that take place once this target has been encountered visually. We measured N2pc components triggered in response to images of real-world target objects that were accompanied by three distractor objects in the same search display (Figure 1). The N2pc is an event-related brain potential (ERP) component that provides a temporally precise index of the covert deployment of spatial attention to targets among distractors in multi-stimulus visual displays (e.g., Luck & Hillyard, 1994; Eimer, 1996; Woodman & Luck, 1999). When a target is presented in the left or right visual field, its attentional selection is reflected by an enhanced negativity at contralateral posterior electrodes (N2pc) that typically starts around 180-200 ms after stimulus onset, and is generated in extrastriate areas of the ventral visual processing stream (Hopf et al., 2000).

In our experiment, each trial run started with a word cue that specified the target object for this run. This word cue was followed by three successive search displays that all contained this target among three distractor objects. Participants' task was to localize the target in each of these three search displays. There were 175 trial runs, and a new target object was specified for each run. Each individual target object only featured in one trial run, and never appeared as a distractor in any other search display. The attentional selection of the target in the 1st display of each trial run had to be guided by an attentional template that was set up in response to the word cue. In contrast, the selection of the 2nd and 3rd target in each run followed the first visual encounter with this target object, and might therefore be controlled by a different search template that specified the visual target properties more comprehensively. If attentional templates for search target objects that are set up in response to verbal cues guide target selection less efficiently than templates that are implemented after a target object has already been seen, this should be reflected in systematic performance and electrophysiological differences between the 1st and the two subsequent search displays. Reaction times (RTs) in response to the initial presentation of the target should be slower than RTs to the 2nd and 3rd appearance of the same target objects. If this RT difference were due to a delay in the allocation of spatial attention to target objects in the 1st display, the N2pc triggered by these target objects should emerge later than the target N2pc or the two subsequent displays. This N2pc delay provides an objective estimate of the time costs associated with the guidance of visual search by verbally as compared to visually cued search templates on early visual-perceptual stages of attentional target selection. If precise attentional templates are implemented gradually in the course of each trial run, N2pc components may also be larger and emerge earlier in response to targets in the final display of each trial run relative to targets in the 2nd display. Alternatively, if a single perceptual encounter with a target object is sufficient to activate an exact target-matching search template, there should be no systematic N2pc differences between the 2nd and 3rd target in each trial run.

Search templates set up in response to verbal target descriptions may guide the allocation of spatial attention more effectively for some target objects than for others. For targets that have a canonical shape or color (e.g., a banana), verbal instructions may be sufficient to set up a precise attentional template, resulting in attentional selection processes that are as efficient as those observed with picture cues. For other visual objects (e.g., bags), which are more varied in terms of their perceptual attributes, word cues may not be sufficient to activate a precise visual representation of the search target. We refer to this as differences in the "*imageability*" of particular objects. This term is often employed in language research to describe participants' self–reported ability to evoke a mental image of an object in response to a word label (e.g., Gilhooly & Logie, 1980). Here, we use imageability to describe differences in the ability of a word cue to consistently trigger target-matching search templates. For a highly imageable target object with invariant visual properties, an attentional template set up in response to a word cue may include a particular mental image of this object, or a set of canonical object features, either of which is likely to provide a close match with the target when it is encountered in a search display. For less imageable search targets with more varied or less canonical visual attributes, templates elicited by word cues are unlikely to precisely match the actual target object or some of its features. Because the efficiency of visual search depends on the match between search templates and target objects, search for targets defined by word cues should differ systematically as a function of their imageability.

Initial evidence for this hypothesis was provided by Castelhano, Pollatsek, & Cave (2008) in an eye tracking study where participants searched for real-world objects that were defined by picture cues or by word cues, and were typical or atypical exemplars of a particular object category. Targets were found faster when search was guided by picture cues, irrespective of target typicality. In contrast, typicality had a strong effect when targets were specified by word cues, with substantially delayed RTs for atypical targets. Interestingly, Castelhano et al. (2008) found that the time between search display onset and the first fixation on the target did not differ between typical and atypical targets in the word cue condition. Based on this observation, these authors concluded that the rapid guidance of attention towards target objects specified by word cues is not affected by the typicality of these targets, and that the performance costs observed for atypical as compared to typical targets are generated at a later object identification stage.

We re-assessed this conclusion, and investigated whether the ability of word cues to facilitate effective template-guided attentional selection processes varies between more and less imageable objects by comparing N2pc components triggered by these objects in the 1st search display in each trial run. Because there is no objective way to determine *a priori* to what degree a particular word cue constrains the visual attributes of a future target object, we employed the RTs measured for the 1st target object in each trial run as a means to separate objects in terms of their imageability. If attentional templates set up by word cues generally provide a better match with more as compared to less imageable target objects, this should be reflected by systematic RT differences when these targets are encountered for the first time immediately after the word cue. We performed a three-way (tertile) split of RTs to the 1st target in each run, separately for each individual participant, and computed ERP waveforms for visual objects that were associated with fast, medium, or slow RTs when they were seen for the first time. If differences in object imageability affect the speed with which these objects can be selected in the 1st display after a word cue, highly imageable target objects should trigger earlier and larger N2pc components than less imageable objects. If a single perceptual encounter was sufficient to implement an efficient attentional template even for objects whose visual properties are only weakly constrained by their word cue, these N2pc differences should be largely eliminated for the 2nd and 3rd display in each trial run. In contrast, if differences in the imageabilty of individual objects primarily affect identification processes that take place after these objects have been selected but not the efficiency of attentional guidance itself (as suggested by Castelhano et al., 2008), there should be no systematic N2pc differences between highly and less imageable target objects in the present study.

In order to attribute any N2pc differences between the three successive search displays in each trial run to differences in the precision of attentional templates, it is important to rule out the possibility that they are instead associated with template-unspecific short-term training effects within each run (i.e., a generic improvement in the efficiency of attentional target selection when search for the same target object is performed for the second or third time). We therefore ran a control experiment that was identical to the main experiment, except that word cues were replaced by picture cues that physically matched the target object for each trial run. Because these picture cues enabled observers to activate a visually precise attentional template prior to the arrival of the 1st search display, there should no longer be any template-related N2pc differences between the three successive displays in each trial run.

Methods

Participants

Fourteen paid volunteers participated in the main experiment (M=31.75 years, SD=8.89, range: 21-50 years, 10 males). All of them had normal or corrected vision, and all were native English speakers. Eight different paid volunteers with normal or corrected vision took part in the control experiment (M=29 years, SD=3.07, range: 27-36 years, 4 males).

Stimuli, Design, and Procedure

The stimuli employed in this experiment were color photographs of real-world objects that were selected from the Boss Normalized stimuli set (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and The Object Databank (Center for the Neural Basis of Cognition, CMU). The stimulus set contained a total of 350 different object images. Object files were pre-processed to generate images of identical size (1.72°x 1.72°). Each object was assigned a specific verbal label that was used as the word cue in the main experiment. To confirm that all objects matched their respective verbal descriptions, we ran an online pilot study with 72 participants (mean age 30 years; range 18-60 years; 26 male). On each trial, a particular object image was shown at fixation, and participants were asked to identify this object by entering free text. Next, the pre-assigned word label for this object was presented, and participants rated the typicality of the object image in relation to its verbal label on a five-point Likert scale (5 - very typical; 1 – not typical at all). Objects were generally rated as highly typical of their label, with a mean typicality score of 4.43 (minimum: 3.14; maximum: 5.0). Because all 350 objects included in our experiment received an above-average typicality score, none of them was removed as a result of this rating study.

During the experiment, stimuli were presented against a white background on a 24-inch LCD monitor with a 100 Hz refresh rate at a viewing distance of 100 cm. A central fixation point was continuously present, and participants were instructed to maintain central fixation throughout each experimental block. Each trial run started with a word cue (1600 ms duration) that specified the target object for this particular run of search displays (Figure 1). 1000 ms after the offset of this cue, the first of three consecutive search arrays was displayed. Each search array contained four images of four different objects in the four quadrants of the visual field at an eccentricity of 2° (measured relative to the centre of each object). Search displays remained visible until a response was recorded. The interval between the search display offset and the onset of the next search display in a run was 1000 ms. The offset of the final display in a given trial run and the onset of the word cue on the subsequent trial run were separated by an interval of 1600 ms.

The experiment contained seven blocks, with 25 trial runs per block, resulting in a total of 175 trial runs. Participants' task was to find the target object specified by the word cue in all three search displays of each trial run, and to report its vertical location (upper versus lower visual hemifield) by pressing one of two vertically arranged response keys with their left or right index finger. All three search displays contained one target object at a randomly determined location among three different distractor objects. Each individual target object was only employed for one trial run, and was never repeated as target or distractor in any other trial run. To implement this constraint, the stimulus set of 350 objects images was divided into two subsets of 175 images. One of these subsets provided the target objects for the 175 trial runs, while the other subset included all distractor objects. For each search display, three different distractor objects were randomly selected from the distractor set. Target and distractor sets were counterbalanced across participants, such that each of the 350 objects included in the stimulus set served as target on one trial run for seven participants.

The control experiment was identical to the main experiment, except that the word cue was replaced by the image of the target object for each trial run. This image was identical to the target image that appeared in the three successive search displays, and was presented at fixation.

EEG recording and data analysis

EEG was DC-recorded from 23 scalp electrodes at standard positions of the extended 10/20 system (500 Hz sampling rate; 40 Hz low-pass filter) against a left-earlobe reference, and re-referenced offline to averaged earlobes. The continuous EEG was segmented from 100 ms before to 700 ms after the onset of a search array, and was averaged relative to a 100 ms pre-stimulus baseline. Trials with artifacts (horizontal EOG exceeding ± 25 µV, vertical EOG exceeding ± 40 μV, all other channels exceeding ±80 μV) were removed prior to analysis. Following artefact rejection, 91% of all trials were retained in the main experiment and 88% in the control experiment. Averaged waveforms were computed for the 1st, 2nd, and 3rd search display in each trial run, separately for displays with a target on the left or right side. N2pc amplitudes were quantified on the basis of ERP mean amplitudes obtained between 200 and 300 ms after search array onset at lateral posterior electrodes PO7 and PO8. Target N2pc onset latencies were compared between task conditions, using the jackknife-based analysis method described by Miller, Patterson, & Ulrich (1998). An absolute amplitude criterion of 1μV was employed to define N2pc onset. For N2pc analyses based on RT tertile splits, EEG epochs were shortened (-100 ms to 500 ms relative to search array onset), to reduce the number of trials eliminated during artefact rejection, and to maintain acceptable signal-to-noise ratios. Bonferroni corrections were applied to pairwise comparisons of experimental effects where appropriate.

Results

Behavioral performance

Mean reaction times (RTs) on trials with correct responses differed between the 1st, 2nd, and 3rd search display within each trial run, $F(2,26)=199.10$, $p<.001$, $\eta^2=.939$. Responses were considerably slower for the 1st search display in each run (733 ms) relative to the 2nd and 3rd display (467 ms and 465 ms, respectively, both $p<.001$). Accuracy was high (97%), and did not differ between the 1st, 2nd, and 3rd search display within each run, $F(2,26)=1.04$, $p=.366$, η^2

=.074.

N2pc components across all target objects

Figure 2 shows grand-averaged ERPs triggered in the 700 ms interval after search array onset at electrodes PO7/8 in response to targets in the 1st, 2nd, and 3rd search display in each trial run. ERP waveforms are shown separately for electrodes contralateral and ipsilateral to the visual field of the target object in each search array. Figure 2 also includes N2pc difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, separately for the 1st, 2nd, and 3rd display in each trial run. Target objects triggered N2pc components in all three search displays, but the N2pc was strongly attenuated and delayed for the 1st display in each trial run relative to the two subsequent search displays.

N2pc mean amplitudes in the 200-300 ms post-stimulus time window were analysed with a repeated measures ANOVA for the factors laterality (electrode contralateral vs. ipsilateral to the target) and serial position (1st versus 2nd versus 3rd display in each trial run). There was a main effect of serial position, $F(2,26)= 22.3$, $p<.001$, $\eta^2=.609$, as ERPs in the N2 time window were generally more positive for the 1st relative to the 2nd and 3rd search display in each trial run (see Figure 2). There was also a main effect of laterality, $F(1,13)= 28.1$, $p<.001$, $\eta^2=.684$, confirming the presence of reliable target N2pc components. Most importantly, an interaction between laterality and serial position, $F(2,26)= 31.5$, $p<.001$, $\eta^2=.708$, suggested that N2pc amplitudes were reduced for the 1st relative to the 2nd and 3rd search display in each trial run. This was confirmed by follow-up analyses of N2pc difference waveforms, which demonstrated significant target N2pc amplitude differences between the 1st and 2nd display, $t(13)=6.67$, $p<.001$, and between the 1st and 3rd display, $t(13)=7.31$, $p<.001$, but no difference between the 2nd and 3rd display, $t(13)<1$. Although the N2pc component was reduced in size for the 1st target presentation, it was reliably present not only in response to the 2nd and 3rd target in each trial run, $t(13)=5.27$ and 7.18, respectively, both $p<.001$, but also for the 1st target presentation, $t(13)= 2.917$, $p=.012$.

The jackknife-based analysis of N2pc latencies with a fixed onset criterion of 1μV revealed a significant effect of serial position, $F_c\ (2,26)=3.54$, $p=.044$, as the onset of the N2pc to target objects in the 1st display (226 ms after display onset) was delayed relative to the target N2pc for the 2nd and 3rd display in each trial run (189 ms and 188 ms, respectively; see Figure 2, bottom right panel). This N2pc onset delay for the 1st relative to the 2nd and 3rd target display was reliable, $t_c(13)=2.41$ and 2.15, respectively, both $p<.05$. There was no N2pc onset latency difference between the 2nd and 3rd display in each run, $t_c(13)<1$.

As can be seen in Figure 2, the attenuated N2pc to target objects in the 1st display during the 200-300 ms time interval was followed by a sustained contralateral negativity at longer post-stimulus latencies, which presumably reflects the latency variability of N2pc components on these trials. This late sustained negativity was much smaller for targets in the 2nd or 3rd display. An analysis of ERP mean amplitudes measured in the 400-700 ms time window revealed an interaction between laterality and serial position, $F(2,26)= 7.4$, $p=.003$, $\eta^2=.362$. Additional analysis confirmed that the late contralateral negativity within this time interval was indeed reliably larger for the 1st display in each trial run relative to the 2nd or 3rd display, $t(13) =4.31$ and 2.57, $p< .001$ and .015, respectively.

N2pc components as a function of target imageablity

Different target objects may vary considerably in their imageability, and this may affect the efficiency of attentional target selection controlled by word cues. To identify target objects with high, intermediate, and low imageability, we performed an RT tertile split, based on response latencies measured for the 1st search display in each trial run that were computed individually for each participant. Mean RTs (averaged across all participants) were 483 ms (±71 ms), 710 ms (±128 ms) 1077 ms (±109ms) for the 1st, 2nd, and 3rd RT tertile. Figure 3 (top panel) shows examples of target objects with high or low imagebility that were that were consistently associated with fast RTs or slow RTs when they were first encountered in a trial run. The results of the RT tertile splits were used to compute target N2pc components separately for objects that triggered fast, medium, or slow RTs upon their initial presentation. Figure 3 (middle panels) shows N2pc difference waveforms obtained for the 500 ms post-stimulus time interval for these three types of objects, separately for their 1st, 2nd, and 3rd presentation within a trial run.

Following their 1st presentation after a word cue, highly imageable objects triggered larger N2pc components than objects with intermediate imageablity. The N2pc appeared to be entirely absent during the 200-300 post-stimulus interval for the least imageable target objects. This was confirmed by an ANOVA with the factors laterality and imageabilty (fast, medium, or slow responses to the 1st display of a particular trial run), which revealed a main effect of laterality, F (1,13)=8.49, $p=.012$, $\eta^2=.395$, and, importantly, a significant interaction between laterality and imageability, F (2,26)= 17.43, $p<.001$, $\eta^2=.573$. Follow-up analyses confirmed the presence of reliable N2pc components for objects with high and intermediate imageability, $t(13)=3.81$ and 3.39, respectively, both $p> .005$, whereas no N2pc was present for the least imageable objects, $t(13)=1.394$, $p=.187$. Target N2pc amplitudes were larger for objects with high versus intermediate imageabilty, $t(13)=2.23$, $p<.05$. These findings demonstrate that differences in the ability of word cues to constrain the expected visual attributes of an upcoming target object can have profound effects on the speed and efficiency of attentional target selection in visual search.

Figure 3 (middle panels) also shows N2pc components to the same three groups of target objects in the 2nd and 3rd display of each trial run, after they had already been encountered in the 1st display. The large N2pc differences observed for their first presentation were now completely eliminated. Analyses of N2pc mean amplitudes with the factors laterality and imageability, revealed main effects of laterality for the 2nd and 3rd display, $F(1,13) = 31.0$ and 57.9, both $p<.001$, $\eta^2 =.704$ and .817, respectively. Critically, there were no longer any interactions between laterality and imageability, both $F(2,26) < 1$, demonstrating that N2pc components of equivalent size were now elicited by all target objects irrespective of their imageability. There were also no reliable N2pc onset latency differences between these objects with high, intermediate, or low imageability for their 2nd and 3rd presentation on each trial run, both $F_c(2,26) = 1.8779$, p=.173, and $F_c(2,26)<1$, respectively.

As highly imageable objects were already associated with fast RTs and large N2pc components on their 1st presentation within a trial run, it is important to determine whether the attentional selection of these objects would still be more efficient after they had been encountered once. Figure 3 (bottom panel) shows N2pc difference waveforms for target objects with fast responses for their 1st presentation, separately for the 1st and 2nd display of a trial run. The N2pc to these objects was triggered reliably earlier when they were encountered for the second time relative to their first presentation (168 ms versus 209 ms post-stimulus, t_c (13)=3.72, p<.008). In line with this observation, mean RTs to these highly imageable target objects were also reliably faster in the 2nd display of a trial run relative to their 1st presentation (436 ms versus 483 ms, t (13) = 6.814, p <.001).

Control Experiment

In this experiment, in which word cues were replaced by picture cues, target RTs were slower for the 1st display in each trial run (503 ms) relative to the 2nd and 3rd display (462 ms and 471 ms), resulting in a main effect of serial position on mean RTs, $F(2,14)=12.57$, $p=.008$, η^2 =.642. Follow-up analyses confirmed that this RT delay for the 1st relative to the 2nd and 3rd presentation of a target object was significant, both $p < .05$. Mean accuracy was 98%, and did not differ between the 1st, 2nd, or 3rd display in each run.

Figure 5 shows contralateral-ipsilateral N2pc difference waveforms obtained in this control experiment in response to target objects in the 1st, 2nd, and 3rd display. In marked contrast to the results obtained in the main experiment (Figure 2, bottom right panel), N2pc amplitudes and onset latencies were unaffected by the serial position of a search display within a trial run, and were now equally large for the 1st presentation of a target object and for the two subsequent target presentations. The analysis of N2pc mean amplitudes revealed a main effect of laterality $F(1,7)= 47.27$, $p<.001$, $\eta^2=.871$, but no interaction between laterality and serial position $F(2,14)<1$. N2pc onset latencies were virtually identical for the 1st, 2nd, and 3rd display in a trial run (179 ms, 179 ms, and 185 ms post-stimulus, respectively, $F_c\ (2,14) < 1$.

Discussion

In real-world contexts, we often search for verbally defined target objects. If search is guided by attentional templates, and if these templates are analog visual representations of search targets, word cues may be less efficient than picture cues in setting up precise search templates. We employed the N2pc component as an electrophysiological marker of attentional target selection to compare the speed of selecting search targets specified by a word cue to the selection of the same targets in subsequent search episodes after these objects have been seen at least once. Our results demonstrate that the guidance of target selection in visual search is often quite inefficient with word cues. RTs were more than 250 ms slower in the 1st search display of each trial run that immediately followed the word cue relative to RTs to targets in the two subsequent search displays. N2pc components in response to the 1st target in each run were also strongly attenuated and delayed relative to the next two targets (Figure 2), demonstrating substantial costs for the speed of attentional target selection when it has to be guided exclusively by a verbal specification of target identity. Across all target objects, the onset delay of the N2pc to the 1st target relative to the 2nd and 3rd target was much smaller (about 30 ms) than the corresponding delay of target RTs, which reflects the variability in the efficiency of attentional guidance by word cues between different target objects. The sustained contralateral negativity beyond the standard N2pc time window for the 1st target in each run

(Figure 2, bottom left panel) suggests that the onset latency of N2pc components elicited by these targets varied substantially as a function of the imageability of individual target objects (see below). If the N2pc is triggered early for some objects and is delayed by a variable amount for others, N2pc amplitudes will be attenuated during the 200-300 ms post-stimulus interval, and a sustained contralateral negativity will emerge at longer latencies.

In contrast to the substantial N2pc and RT differences between the 1^{st} and 2^{nd} target in each trial run, there were no performance or ERP differences between the 2^{nd} and 3^{rd} presentation of a particular target object. RTs as well as target N2pc amplitudes and onset latencies were essentially the same for these two search displays (Figure 3). These findings demonstrate that a single visual presentation of a particular target object is sufficient to establish a precise attentional template, and that there are no additional benefits for target selection in subsequent search episodes.

The results from the control experiment demonstrated that the performance and N2pc differences observed between the 1^{st} and subsequent presentations of a target were not simply due to observers' increased practice in selecting a particular target object during each trial run. In this control experiment, where word cues were replaced by an exact image of the target for each trial run, all N2pc amplitude or onset latency differences between the 1^{st}, 2^{nd}, and 3^{rd} search display were eliminated (Figure 4). This demonstrates that when a perceptually precise attentional template can be implemented prior to the 1^{st} search display in each trial run, target selection already operates efficiently for this display, and shows no further improvement for subsequent search episodes with the same target object. It should be noted that there was a small but reliable RT cost of about 40 ms for the 1^{st} display relative to the 2^{nd} and 3^{rd} display in each trial run in this control experiment. The absence of any corresponding N2pc latency differences strongly suggests that this RT difference was generated at stages that follow the template-guided selection of target objects, such as the identification of a selected visual object as the target (e.g., Castelhano et al., 2008; Eimer, 2014) and the activation of a corresponding response. For example, it is likely that a manual response to a particular target object will be selected and executed faster when the same response to the same object has already been activated for a preceding search display.

The ability of word cues to trigger visually precise search templates may differ as a function of the imageability of target objects. Because each of the 350 objects used in this study only served as target on a single trial for seven participants, determining their imageability in an item-specific fashion on the basis of the RTs measured on these seven trials is likely to yield a relatively low signal-to-noise ratio. We therefore chose a different approach, and performed an RT-based tertile split and computed separate N2pc components for target objects that were associated with fast, medium, or slow RTs when they were first encountered in each trial run. Because this tertile split was based on the overall RT distributions across all trials for individual participants, it could in principle have been affected not just by target imageability, but also by the similarity of target and distractor features on single trials (although distractors were randomly selected on each trial). The fact that the classification of individual objects in terms their imageability obtained with this method and the item-specific classification based on RTs of seven trials were closely correlated (r=.747; p <.001) demonstrated that these classifications tended to be consistent across participants and target objects.

In the 1st display of each trial run, N2pc components in the 200-300 ms post-stimulus time interval were largest when target objects were highly imageable, and entirely absent for the least imageable objects (Figure 3). The absence of any early N2pc for this latter group of objects suggests that they were selected much later, and beyond the 500 ms post-stimulus analysis interval that was employed for these tertile split analyses. These N2pc results demonstrate that there are large differences in the ability of word cues to constrain the perceptual properties of real-world search targets, and that these differences have important consequences for the efficiency of attentional target selection in visual search. For some objects, a verbal label is sufficient to form an attentional template that matches their perceptual attributes, and these objects can then be selected efficiently. For other objects, word cues do not facilitate the implementation of a precise target-matching attentional template, resulting in inefficient target selection. Importantly, these N2pc differences between individual target objects were only observed for their 1st presentation within each trial run, but were eliminated when the same objects reappeared for the second and third time (Figure 3, middle panels). This demonstrates that once an object has been visually perceived, a precise attentional template can be formed, regardless of whether a word cue had previously been effective or ineffective in facilitating an attentional template for this object. Even the most imageable objects that were associated with fast RTs and the large N2pc components when they appeared immediately after a word cue were selected more efficiently once they had been encountered visually, as reflected by faster RTs and shorter-latency N2pc components during their 2nd presentation within a trial run (Figure 3, bottom panel). This finding suggests a generic limitation in the ability of verbal descriptions to facilitate the formation of precise attentional templates, even for highly imageable target objects (see also Schmidt & Zelinsky, 2009).

What is the nature of the search templates that are activated in response to verbal descriptions of real-world target objects as were used in the present study, and how is the efficiency of template-guided search affected by differences in the imageability of particular target objects? An attentional template may be an analog visual representation of a whole object or a set of independent features that are expected to match the visual features of the anticipated target object (see Eimer & Grubert, 2014, for a dissociation between feature-based and object-based attentional control in the selection of targets defined by a conjunction of simple features, and Evans & Treisman, 2005, for a distinction between the object-based and feature-based detection of targets in natural visual scenes). For highly imageable objects with invariant visual properties, word cues should be able to trigger object or feature templates that closely match the perceptual attributes of the actual target objects, resulting in their efficient selection. When less imageable objects with more variable properties are specified by word cues, participants might set up one particular object representation which is less likely to match the target in the 1st display, a set of possible target features that may or may not be shared by the actual target object, or may not activate a visual search template at all. In all three of these scenarios, template-guided target selection will be less efficient relative to more imageable objects. The observation that even the most imageable objects were selected more efficiently once they had been encountered visually could be linked to the difference between feature-based and object-based search templates. Word cues may generally only be able to activate representations of one or more target-defining features, whereas a full analog object search template can only been implemented if this object has been seen at least once.

The general superiority of picture cues over word cues, and the effects of object imageability on the efficiency of attentional target selection following word cues both highlight the importance of a close perceptual match between an attentional template and target objects during attentional guidance in visual search. The effective guidance of spatial attention towards particular real-world targets depends on the activation of visual search templates that match the perceptual properties of these target objects. However, this may not be the case for other types of visual search tasks. In a recent set of ERP studies (Wu et al., 2013; Nako, Wu, & Eimer, 2014a; Nako, Wu, Smith, & Eimer, 2014b), we employed N2pc components to assess the efficiency of category-based attentional selection in visual search. When participants searched for category-defined alphanumerical items (e.g., any letter among digits; Wu et al., 2013; Nako et al., 2014a) or real-world objects (e.g., any kitchen object among items of clothing: Nako et al., 2014b), targets that matched the currently relevant category triggered early N2pc components that emerged around 180 ms (alphanumerical search) or 240 ms post-stimulus (search for category-defined real-world objects), demonstrating that target selection can be fast and efficient even when it cannot be based on an attentional template that specifies particular visual attributes of a target object. This suggests that search templates may not always be pictorial representations of visual target attributes, but can also represent more abstract target-defining properties. Which type of template is active may depend on the selection demands of a particular search task. When targets are defined at the category level, search templates may represent abstract target categories. When participants search for a specific target object, as in the present study, target selection may be exclusively guided by representations of the visual object features. If this was the case, tasks that encourage category-based selection and tasks that emphasize perceptual target attributes should produce qualitatively distinct patterns of attentional guidance, even when search displays are physically identical. This possibility will need to be addressed in future research.

The present results demonstrate that the imageability of individual target objects strongly affects the efficiency of attentional guidance and target selection during visual search when target identity is specified by word cues. This conclusion appears to be inconsistent with the results from the eye tracking study by Castelhano et al. (2008). These authors found that the time from search display onset to the first fixation on the target did not differ between typical and atypical target objects in a word cue condition, and concluded that performance costs during search for atypical targets mainly originate at a post-selection object identification stage. It is likely that Castelhano et al. (2008) did not find effects of target typicality on attentional guidance because in their experiment, distractor objects always shared one or more features with the target, resulting in very inefficient search. This was reflected by slow RTs (above 1500 ms in the word cue condition), and by the fact that on most trials, there were several eye movements to distractor objects before the target was fixated. If the search templates that can be activated in response to word cues are always representations of one or more specific target features rather than visual representations of whole target objects (as suggested above), such feature-based templates may not be useful for the rapid attentional guidance of target selection when target and distractor objects share features, as in the Castelhano et al. (2008) study.

Overall, the present study has provided new electrophysiological evidence that during search for real-world objects, early perceptual stages of attentional target selection are strongly delayed when search targets are specified verbally as compared to search for visually defined targets. Although the ability to implement an effective search template in response to word cues varies greatly between more and less imageable target objects, a single visual presentation of a particular object is sufficient to activate a precise attentional template.

References

Brodeur, M.B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5, e10773.

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. Journal of Experimental Psychology: Human Perception and Performance, 34, 660.

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

Dombrowe, I., Donk, M., & Olivers, C. N. (2011). The costs of switching attentional sets. *Attention, Perception, & Psychophysics*, 73, 2481-2488.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. Electroencephalography and *Clinical Neurophysiology*, 99, 225–234.

Eimer, M. (2014). The neural basis of selective attention in visual search. *Trends in Cognitive*

Neuroscience. Online version, doi: http://dx.doi.org/10.1016/j.tics.2014.05.005.

Eimer, M. & Grubert, A. (2014). The gradual emergence of spatially selective target processing in visual search: from feature-specific to object-based attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, Jul 7 , 2014, No Pagination Specified. doi: 10.1037/a0037387.

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*, 1476.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*, 395-427.

Hopf, J.-M., Luck, S. J., Girelli, M., Hagner, T., Mangun, G. R., Scheich, H., & Heinze, H.-J. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex*, 10, 1233-1241.

James, W. (1890). The Principles of Psychology, Harvard University Press.

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: a computational approach. *Psychological Review*, 94, 148.

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, 129, 723.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31, 291-308.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, 35, 99-115.

Nako, R., Wu, R., and Eimer, M. (2014a). Rapid guidance of visual search by object categories. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 50-60*.*

Nako, R., Wu, R., Smith, T. J., & Eimer, M. (2014b). Item and Category-Based Attentional Control During Search for Real-World Objects: Can You Find the Pants Among the Pans? *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1283-1288.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not*. Trends in Cognitive Sciences*, 15, 327-334.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and brain sciences*, 25, 157-182.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology*, 62, 1904-1914.

Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*, 8.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202-238.

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 483-502.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 495-501.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, 44, 1411-1426.

Wolfe, J. M. (2007). Guided search 4.0. *Integrated models of cognitive systems*, 99-119.

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, 400, 867-869.

Wu, R., Scerif, G., Aslin, R. N., Smith, T. J., Nako, R., & Eimer, M. (2013). Searching for something familiar or novel: Top-down attentional selection of specific items or object categories. *Journal of Cognitive Neuroscience,* 25*,* 719-729.

Acknowledgements

The authors thank Monica Castelhano, Anna Grubert, and an anonymous reviewer for helpful comments on previous versions of this manuscript, and Susan Nicholas for technical support. This work was supported by the Economic and Social Research Council, United Kingdom (ES/K006142/1), and the BIAL Foundation (#224/12).

Figure Legends

Figure 1. An example of a trial run in the main experiment. Each trial started with a word cue that specified the target object for this run. The word cue was followed by three successive search arrays that all contained the target object among three different distractor objects. In the control experiment, word cues were replaced by an image of the target object.

Figure 2. Grand-average ERP waveforms elicited in response to the 1^{st}, 2^{nd}, and 3^{rd} display in each trial run at posterior electrodes PO7/8 contralateral and ipsilateral to a target object. The bottom left panel shows N2pc difference waveforms obtained by subtracting ipsilateral from contralateral ERPs for the 1^{st}, 2^{nd}, and 3^{rd} display in each trial run.

Figure 3. Top panel: The six most imageable and the six least imageable target objects and their associated word cues. These objects were consistently associated with fast responses (mean RT across all participants: 435 ms) or slow responses (mean RT: 1153 ms) when they first appeared in a trial run immediately after the word cue. Middle Panel : N2pc difference waveforms obtained by subtracting ipsilateral from contralateral ERPs for target objects with high, intermediate, or low imageability, shown separately for the 1^{st}, 2^{nd}, and 3^{rd} presentation of the same target objects within each trial run. Bottom Panel: N2pc difference waveforms in response to the most imageable target objects for their 1^{st} and 2^{nd} presentation within each trial run.

Figure 4. N2pc results obtained in a control experiment where word cues were replaced by picture cues. N2pc difference waveforms obtained by subtracting ipsilateral from contralateral ERPs at lateral posterior electrodes PO7/8 are shown separately for the 1^{st}, 2^{nd}, and 3^{rd} display in each trial run.

Figure 1

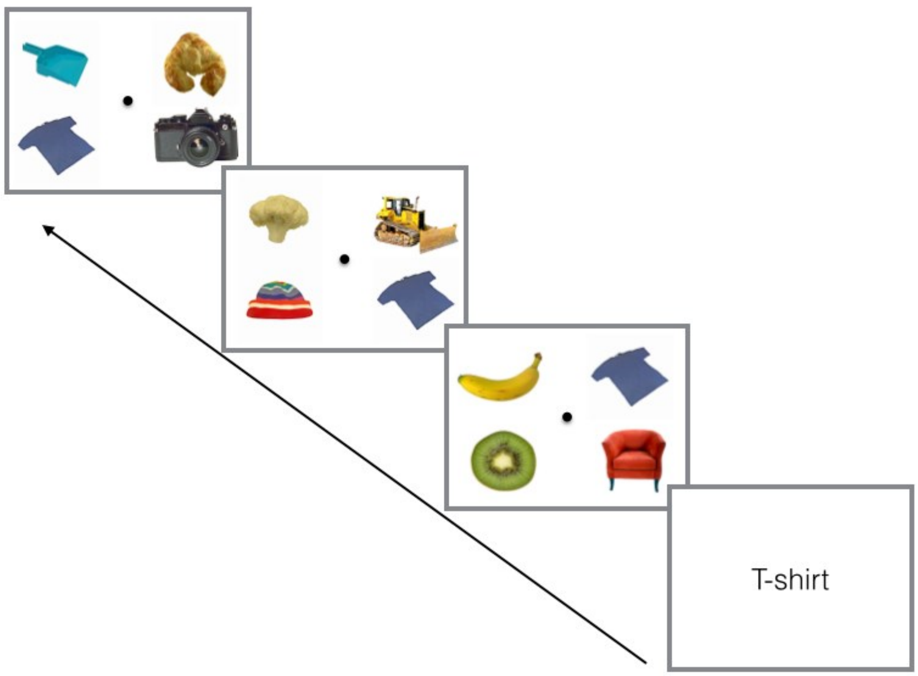


Figure 2

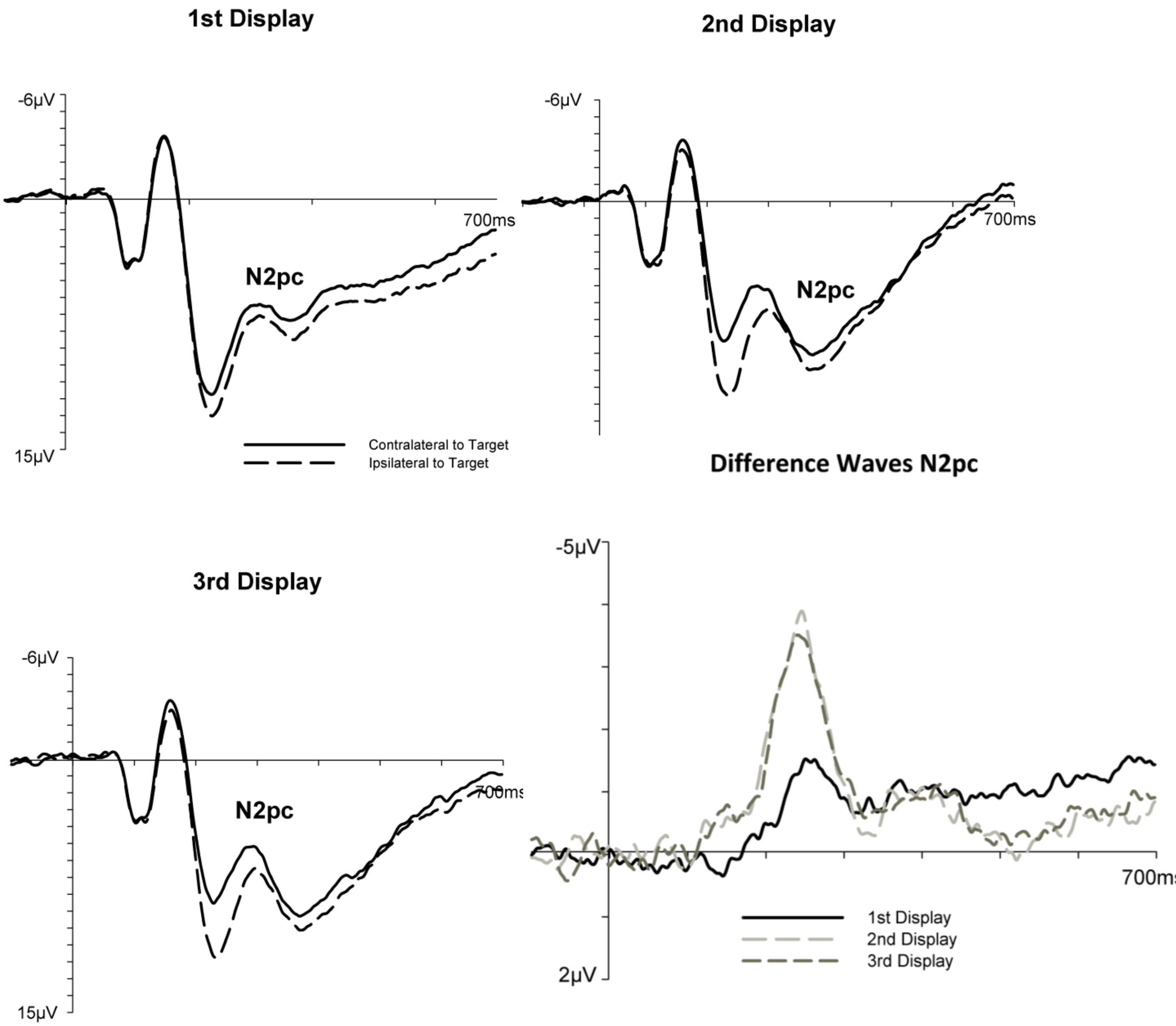

1st Display
-6µV
700ms
N2pc
15µV
Contralateral to Target
Ipsilateral to Target
2nd Display
-6µV
700ms
N2pc
3rd Display
-6µV
700ms
N2pc
15µV
Difference Waves N2pc
-5µV
700ms
1st Display
2nd Display
3rd Display
2µV

Figure 3

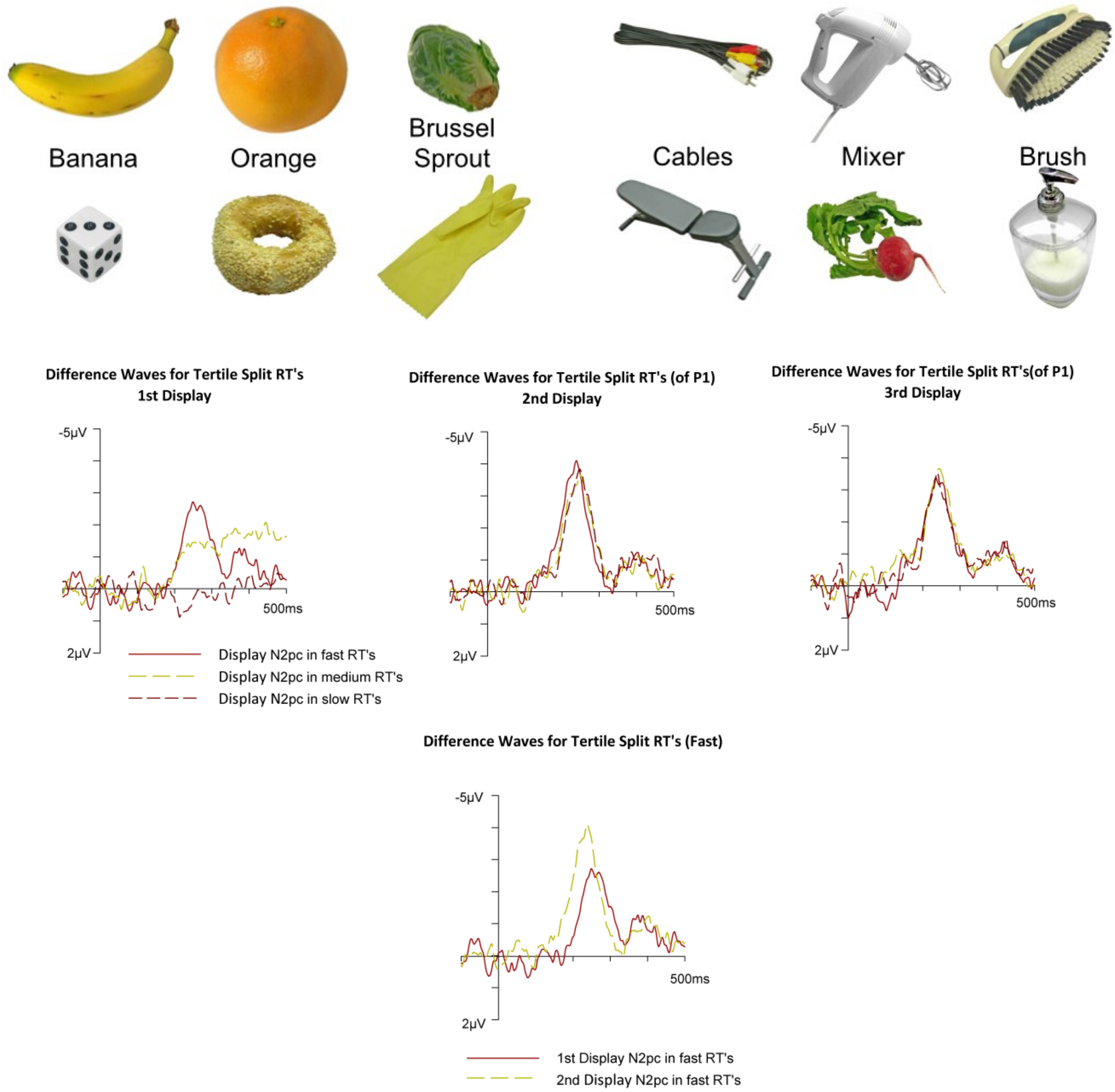
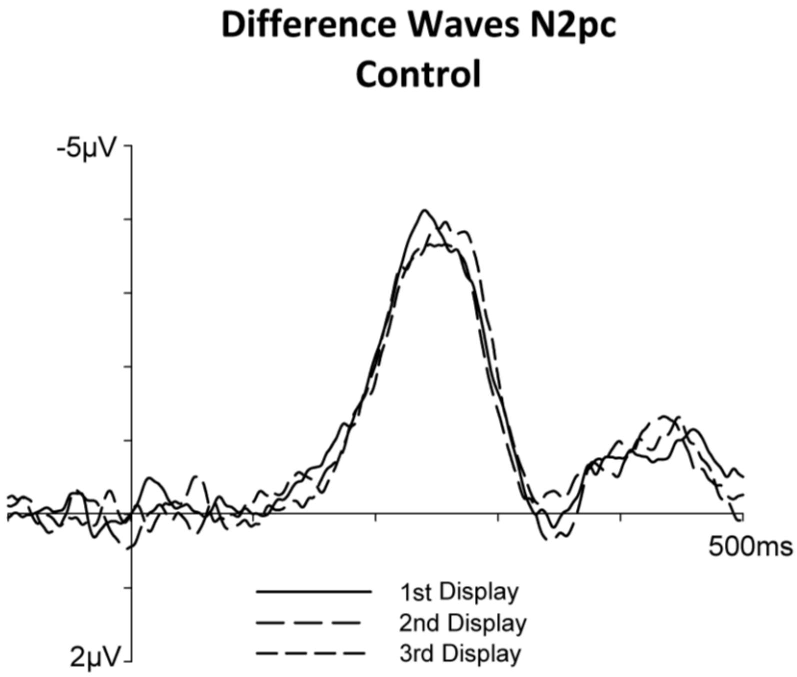

Banana
Orange
Brussel Sprout
Cables
Mixer
Brush
Difference Waves for Tertile Split RT's
1st Display
-5μV
2μV
500ms
Display N2pc in fast RT's
Display N2pc in medium RT's
Display N2pc in slow RT's
Difference Waves for Tertile Split RT's (of P1)
2nd Display
-5μV
2μV
500ms
Difference Waves for Tertile Split RT's(of P1)
3rd Display
-5μV
2μV
500ms
Difference Waves for Tertile Split RT's (Fast)
-5μV
2μV
500ms
1st Display N2pc in fast RT's
2nd Display N2pc in fast RT's

Figure 4


Difference Waves N2pc
Control
-5µV
2µV
500ms
1st Display
2nd Display
3rd Display