Complex naturalistic stimuli capture attention automatically when maintained in working memory

Natalia Rutkowska, Łucja Doradzińska, Michał Bola

Laboratory of Brain Imaging, Nencki Institute of Experimental Biology, 3 Pasteur Street, 02-093 Warsaw, Poland

Corresponding author:

Michał Bola, PhD

Email: m.bola@nencki.edu.pl

Laboratory of Brain Imaging

Nencki Institute of Experimental Biology

3 Pasteur Street

02-093 Warsaw, Poland

Conflict of interest: The authors declare no competing interests.

Acknowledgements: This study was funded by a National Science Center Poland (grant number 2018/29/B/HS6/02152).

Abstract

Recent studies suggest that a stimulus actively maintained in working memory (WM) automatically captures visual attention when subsequently perceived. However, such a WM guidance effect has been so far observed only for stimuli defined by simple features, such as colour or orientation. Here we investigated whether the effect occurs also for naturalistic stimuli, whose identity is defined by multiple features and relations among them, specifically for faces and houses. The experiment comprised two conditions: a WM condition and a mere exposure condition. Subjects (N = 28) either memorized or merely saw a template stimulus, and then performed several dot-probe trials, in which pairs of stimuli were presented laterally as distractors (memorized or seen on the one side, control stimulus on the other) and followed by a target dot. We found that response times were faster when targets followed a memorized stimulus than when they followed a control one, and that the memory-matching stimuli evoked the N2pc ERP component. Further, in an exploratory analysis we found electrophysiological evidence for an early (100-200 ms post stimulus) prioritization specific to the memorized faces. Importantly, neither the RT, nor the electrophysiological effects were observed in the mere exposure condition. In conclusion, showing an attention capture by complex stimuli maintained in WM contributes to identifying the mechanism of WM-based attentional selection and provides novel evidence that complex stimuli can, in principle, guide attention automatically. Finally, we interpret the early prioritization of the memorized faces as evidence supporting the sensory recruitment theory of WM.

Keywords working memory, naturalistic stimuli, face perception, attention capture, N2pc

Introduction

Contemporary theories of memory emphasize its role in the prospective guidance of perception and action (Nobre and Stokes, 2019). Particularly, the working memory (WM) system is currently recognized as the key component of the pro-active top-down selection mechanism (Desimone and Duncan, 1995; Nobre and Stokes, 2019). While WM plays an important role in the volitional control of attention, it can also influence attentional selection in an involuntary way. Specifically, a stimulus which is encoded and actively maintained in visual WM automatically attracts attention upon a subsequent presentation (Olivers et al., 2006; Soto et al., 2005, Soto et al., 2007, 2009; review: Soto et al., 2008). This effect has been revealed, first, by visual search experiments, showing increased search times when a WM-maintained stimulus appeared in the search array as a distractor. Second, by eye-tracking studies indicating that eye-movements are attracted by visual input matching the WM content (Hollingworth et al., 2013; Schneegans et al., 2014; Silvis and Van der Stigchel, 2014). Third, by electrophysiological experiments which revealed that the WM-maintained stimuli evoke the N2pc

component, a classic index of covert attention shifts (Carlisle and Woodman, 2013; Kumar et al., 2009). Importantly, such a WM-based attention capture effect is not a form of priming, as it is observed only when a stimulus is memorized but not when it is merely seen; and is considered automatic and involuntary, as it occurs even when detrimental to the task performance (review: Olivers et al., 2011).

The automatic guidance of attention from WM has been so far demonstrated by studies using simple stimuli, defined either by colour or orientation (review: Soto et al., 2008). Stimuli varying on a single dimension of one basic feature are most effective in guiding bottom-up attention, as they can be processed pre-attentively and result in a pop-out search (Wolfe & Horowitz, 2017). Therefore, a question arises whether naturalistic stimuli, whose identity is typically defined by multiple features and relations among them, can cause a similar WM-based attention guidance effect. The general potential of such complex stimuli to guide attention automatically, in either bottom-up or top-down manner, is a matter of ongoing debate (Wolfe & Horowitz, 2017), and thus revealing whether they can benefit from the WM-based attention mechanism will provide an important piece of evidence. Importantly, using complex stimuli will also contribute to a better understanding of the WM-based attention guidance mechanism, as complex stimuli are difficult to define semantically and thus require maintaining a predominantly visual representation, whereas simple ones used previously could have been easily encoded as semantic representations (e.g. "a red triangle"; see: Soto and Humphreys, 2007). Finally, addressing whether complex stimuli might guide attention from WM will open further avenues to investigate the effect in more ecological conditions.

Here we investigated whether two types of complex naturalistic stimuli - images of faces and houses - capture attention automatically when maintained in visual WM. Further, using electrophysiological data we aimed to track the time-course of neuronal activity involved in matching the perceptual input and the WM representation. Faces and houses were chosen as both stimuli types are defined by multiple features and thought to exhibit similar levels of complexity (Filliter et al., 2016). Importantly, investigating two classes of naturalistic stimuli is critical for establishing generalizability of the effect of interest. Therefore, we designed an experiment consisting of multiple blocks, in each a template stimulus (faces or houses) was either memorized for later recollection (WM condition) or merely seen without the need to memorize (exposure condition). Next, in each block participants performed a sequence of dot-probe trials, in which the template and a control stimulus were presented as task-irrelevant distractors (MacLeod et al., 1986). We hypothesized that in the WM condition the templates will involuntarily attract attention, as indicated by an RT effect and the N2 posterior contralateral (N2pc) ERP component, but we did not expect to observe an attention capture in the mere exposure condition.

Methods

Subjects

We analyzed data of 28 subjects (18 females, mean age = 24.2, SD = 2.59 yrs., range: 19-28, 2 left-handed). They all declared normal or corrected-to-normal vision and no history of mental or neurological disorders. Data of 6 additional subjects were collected, but they were excluded from the analysis: 4 subjects due to the technical problems during an EEG recording procedure; 1 subject did not comply with the dot-probe task instruction; and 1 subject due to insufficient number of epochs remaining after EEG signal pre-processing (detailed criteria are described in the "EEG recording and analysis" section).

The study was conducted with the approval of the human ethics committee of the SWPS University of Social Sciences and Humanities (Warsaw, Poland). All subjects provided written informed consent and received monetary compensation for their time (100 PLN= c.a. 25 EUR).

Stimuli

Two sets of stimuli were used. First, 60 pictures[1] of faces with neutral expression (30 male, 30 female; all Caucasian) selected from the Karolinska Directed Emotional Faces stimulus set (KDEF; Lundqvist et al., 1998). From their original format, the face photographs were converted to grayscale using the Gnu Image Manipulation Program (GIMP; available at http://www.gimp.org/). Second, 60 pictures[2] of houses from the DalHouses stimulus set (Filliter et al., 2018). All houses' images were originally in grayscale, presented on a white background, thus no modifications were introduced.

Procedure

The experimental procedure was written in the Presentation software (Neurobehavioral Systems, Albany, CA, USA) and presented on a FlexScan EV-2450 (Hakusan, Ishikawa, Japan) screen through an Intel Core i3 computer. Participants were sat comfortably in a dimly lit room with a viewing distance of 70 cm, which was maintained by a chinrest.

The procedure started with a display providing subjects with the task instructions and information about the trial structure. The procedure consisted of two tasks: a working-memory (WM) task and a mere exposure task; and involved two stimuli types: faces and houses. Thus, there were four conditions: a face WM condition; a house WM condition; a face exposure condition; and a house exposure conditions; which were presented to participants as separate blocks, the order of which was random.

Each condition was further sub-divided into 32 memory or exposure blocks. Each of such blocks started with a central presentation of a template stimulus for 5000 ms (**Fig. 1**). The instruction - either "Memorize this picture" (in the WM condition) or "Take a look at this picture" (in the exposure condition) - was displayed above the image. The template stimulus and a control stimulus were randomly chosen from the pool of all houses or faces before each memory/exposure block. Additionally, in the face conditions the face stimuli were gender-matched, i.e. female and male

template images were paired only with, respectively, female and male control images. In both, a face WM condition and a face exposure condition female faces were used in half of the blocks, and male faces in the other half. All stimuli were presented against a black background. The face images subtended 6.1° x 8.5° of the visual angle, while hause images varied in size and subtended from 4.7° to 7.2° × 6.3° of the visual angle.

After the display of the template stimulus, a sequence of dot-probe trials was presented. Each dot-probe trial started with a fixation cross (subtending 0.4° x 0.4° of the visual angle) displayed in the centre of the screen. The fixation cross remained on-screen throughout the trial. After 1000 ms a pair of stimuli were presented bilaterally for 200 ms - the template stimulus on one side and control stimulus on the other. Face stimuli were presented with their inner edge 3.6° left and right from the fixation cross, while house stimuli with the inner edge from 3.1° to 4.3° left and right from the fixation cross. Next, a target asteriks subtending 0.3° x 0.3° of the visual angle was presented for 150 ms in the location of the centre of either the template stimulus (congruent trial) or the control one (incongruent trial). Participants were instructed to maintain their gaze on the centrally presented fixation cross, ignore the laterally appearing stimuli, and indicate the side of the target dot presentation (left or right) by pressing one of two buttons using index fingers of their left or right hand. Subjects were asked to respond as quickly and accurately as possible. The response time to the target dot was limited to 3000 ms and the next trial started immediately after the manual response.

In the WM conditions, subjects were given a memory test at the end of each block (i.e. after completing the dot-probe trials sequence). Either a template image or a different image (neither a template, nor a control stimulus) was presented centrally. Subjects had to indicate whether the presented stimulus is the same or different from the one they were previously asked to memorize. Above the image a question "Is this the same image?" was displayed. In half of the WM blocks the stimulus was the template image (correct answer "yes"), and in the other half it was a different image (correct answer "no"). The answer "yes" was displayed in the left corner of the screen and the answer "no" in the right corner. Participants responded by pressing one of two buttons (left for "yes", right for "no"). The response time in the memory test was limited to 5000 ms and the next trial started immediately after the manual response. In the mere exposure condition participants were not tested for stimulus recognition, but immediately after completion of the dot-probe sequence the next exposure block started.

Each condition comprised 32 WM/exposure blocks. Within each condition half of the WM/exposure blocks comprised 4 dot-probe trials, and the other half comprised 8 dot-probe trials (the order of blocks was random). The number of trials varied in order to prevent participants from expecting the exact moment of a memory test, and thus encourage them to maintain the WM active throughout the block. Within each dot-probe sequence the template stimulus was presented on the left side in half of the dot-probe trials, and on the right side in the other half. Further, half of the dot-probe trials were congruent and half were incongruent. The order of trials within each sequence was

random. In total 192 dot-probe trials were presented per condition. Participants had a self-paced break five times per condition.

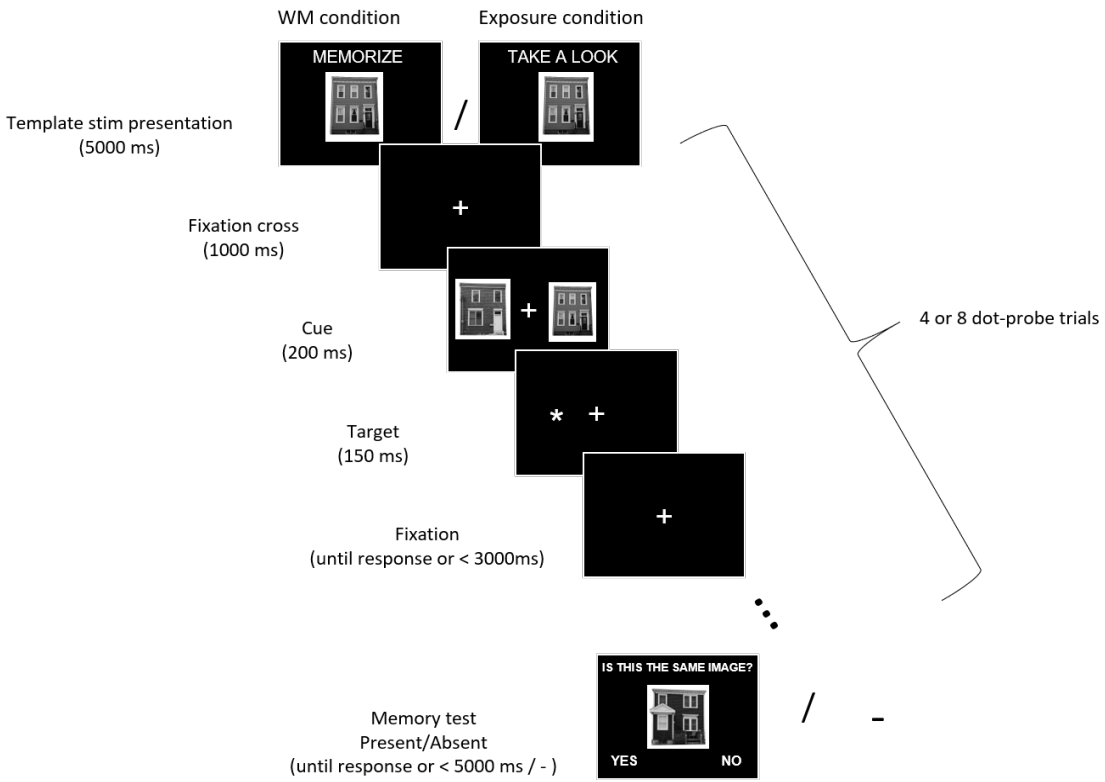

Figure 1. Experimental procedure. Subjects were asked to memorize (WM condition) or just look (exposure condition) at the presented stimulus (either a face or a house). Then they performed 4 or 8 dot-probe trials in which their task was to indicate with a button press the location of a target dot (left or right). Dots were preceded by pairs of stimuli, which subjects were supposed to ignore. A memorized/seen stimulus was presented on one side, and a control one on the other side. In the WM condition subjects completed a memory test, in which they had to decide if the presented stimulus is the one they were instructed to memorize (in the exposure condition there was no memory test).

Analysis of behavioral data

All analyses of behavioural data were conducted using custom-made Python scripts. Therefore, analysis of the dot-probe task data was focused on establishing whether reaction times (RT) of manual responses to the target dot differ between two types of trials: those in which the dot was presented on the same side as the potentially attention-grabbing stimulus (memorized/seen face or house) and those in which the dot was presented on the neutral stimulus side (control face or house). Reaction times were calculated only for the correct responses. For the WM condition the accuracy of memorizing a template stimulus was calculated as a percentage of correct answers in the memory test.

EEG recording and analysis

During the experiment, EEG signal was recorded with 64 Ag-AgCl electrically shielded electrodes mounted on an elastic cap (ActiCAP, Munich, Germany) and positioned according to the extended 10–20 system. Vertical (VEOG) and horizontal (HEOG) electrooculograms were recorded using bipolar electrodes placed at the supra- and sub-orbit of the right eye and at the external canthi. Electrode impedances were kept below 10 kΩ. The data were amplified using a 128-channel amplifier (QuickAmp, Brain Products, Enschede, Netherlands) and digitized with BrainVisionRecorder® software (Brain Products, Munich, Germany) at a 500 Hz sampling rate. The EEG signal was recorded against an average of all channels calculated by the amplifier hardware.

EEG and EOG data were analyzed using EEGlab 14 functions and Matlab 2016b. First, all signals were filtered using high-pass (0.5 Hz) and low-pass (45 Hz) Butterworth IIR filter (filter order = 2; Matlab functions: *butter* and *filtfilt*). Then data were re-referenced to the average of signals recorded from left and right earlobes, and down-sampled to 250 Hz. EEG and EOG data were analyzed using EEGlab 14 functions and Matlab 2016b. First, all signals were filtered using high-pass (0.5 Hz) and low-pass (45 Hz) Butterworth IIR filter (filter order = 2; Matlab functions: *butter* and *filtfilt*). Then data were re-referenced to the average of signals recorded from left and right earlobes, and down-sampled to 250 Hz.

All data were divided into 768 dot-probe epochs (192 epochs per condition; [-200, 1200] ms with respect to the faces/houses images onset) and the epochs were baseline-corrected by subtracting the mean of the pre-stimulus period (i.e. [-200, 0 ms]). Further, epochs were rejected based on the following criteria (all values Mean±SEM): i) when there was no manual response to the target dots until 1.2 s after the onset (18.9±6.0; range [0, 120] epochs per subject); ii) when activity of the HEOG electrode in the time-window [-200, 500] ms exceeded -40 or 40 uV (105.5±19.1; range [11, 352] epochs per subject); iii) when activity of the P7 or P8 electrode in the time-window [-200, 600] ms exceeded -80 or 80 uV (none of the epochs rejected). Thus, after applying the described criteria the average number of analyzed epochs per subject was: 643.6 ±22.0; range [355, 752].

A subject was excluded if the number of epochs in any condition was < 60. This criterion resulted in excluding 1 subject out of 29 (but additional 5 subjects were excluded due to other criteria, as described in the "Subjects" section). The numbers of epochs provided above were calculated based on the final sample of 28 subjects.

Next, each EEG-EOG data-set was decomposed into 50 components using Independent Component Analysis as implemented in the EEGlab *pop_runica* function. To remove residual oculographic artefacts from the data the following procedure was used: time-course of each component was correlated with time-courses of HEOG and VEOG electrodes and in case the Spearman correlation coefficient exceeded -0.3 or 0.3 a component was subtracted from the data. Using this procedure 3.0±0.2 components (range [1, 6]) per subject were removed.

After applying the described preprocessing steps, data were divided with respect to the condition and presentation side of the template stimulus. To calculate the N2pc component we used the P8 and P7 electrodes. Specifically, when template stimulus was presented on the left side, P8 was the contralateral electrode and P7 was the ipsilateral electrode. When template stimulus was presented on the right side, P7 was the contralateral electrode and P8 was the ipsilateral electrode. For each condition contralateral and ipsilateral signals were first concatenated and then averaged, to obtain contralateral and ipsilateral waveforms. Then for each subject and condition, the difference between the contralateral and ipsilateral waveforms was calculated. The 200-400 ms temporal window was used for the N2pc analysis, which is largely consistent with previous studies (e.g. Woodman et al., 2009; Reutter et al., 2017, Wójcik et al., 2019). Further, based on the visual inspection of the obtained ERP waveforms a 100-200 ms window was analysed in an exploratory way.

Statistical analysis

All statistical analyses were conducted in the JASP software and cross-checked with Statcheck (http://statcheck.io/index.php). The values are reported as Mean±SEM, unless stated otherwise.

Bayes Factor (BF_{10}) was used as the primary statistical measure. The main reason for choosing BF_{10} was that, unlike the classic frequentist statistics, BF_{10} evaluates how strongly both alternative and null hypotheses are supported by data. Specifically, BF_{10} is a ratio of the probability (or likelihood) of observing the data given the alternative hypothesis is true to the probability of observing the data given the null hypothesis is true. Thus, in our particular case, BF_{10} allows providing further evidence either in favour or against attention capture by template stimuli.

In all Bayesian tests, the medium prior scale (Cauchy scale 0.707) was used. In the results section we provide interpretations of the BF_{10} according to Wagenmaker et al. (2018), with $0.33 < BF < 3$ indicating inconclusive (anecdotal) evidence. Additionally, for each comparison we also provide results of a frequentist test to complement BF_{10}. Data distribution was first tested with the Shapiro-Wilk test, and a t-test was used when the distribution was Gaussian, or a nonparametric Wilcoxon test when it deviated from normality.

To test for the presence of the dot-probe task effects (RT and N2pc) one-tailed (directional) t-tests were used, with the alternative hypothesis of shorter RT in congruent than incongruent dot-probe trials, and values of the difference (i.e. ipsi - contra) waveform < 0. The early 100-200 ms time window was analyzed in an exploratory way using a two-tailed test. Two-tailed test was also used for comparison of the memory test accuracy scores between faces and houses conditions.

Data availability

The data used in the statistical analysis can be accessed from the OSF (https://osf.io/9rc4j/). Raw EEG data, and scripts used for analysis presentation of the procedure will be shared by authors per request.

Results

Memory accuracy

We first tested whether in the WM conditions participants were actively maintaining the template stimulus in working memory (WM). We found high accuracy values for both faces (97.2±0.7) and houses (94.2±1.6), indicating good performance. Comparing the WM accuracy between face and houses we found anecdotal evidence suggesting that faces were better memorized (BF_{10} = 1.54; W = 172.500, p = 0.010). Thus, our data suggest better WM performance when faces are memorized, which is in line with previous studies (Curby et al., 2009).

Dot-probe task - behavioral results

In the dot-probe task subjects exhibited ceiling-level accuracy in indicating the side of the dot presentation, with the percentage of correct responses being: 97±0.6% in the house WM condition; 97±0.6% in the face WM condition; 97±0.8% in the house exposure condition; 98±0.5% in the face exposure condition. Therefore, the analysis focused on the reaction times (RT) of responses to the target dot. Specifically, we investigated whether RT was shorter when the target followed a potentially attention-grabbing template stimulus (i.e. congruent trials), in comparison to trials when it followed a control stimulus (i.e. incongruent trials). In the houses WM condition we found strong evidence that RT was shorter in congruent trials (350.6±10.13 vs. 368.1±14.15 ms; BF_{10} = 10.62; W = 60.00, p < 0.001), while in the faces WM we found extreme evidence (353.5±7.08 vs. 368.9±6.34 ms; BF_{10} = 230; t(27) = -4.21, p < 0.001). In the exposure conditions Bayes factor indicates moderate evidence for the null hypothesis (i.e. no RT difference) for houses (352.4±12.88 vs. 353.2±11.89 ms; BF_{10} = 0.25; t(27) =- 0.25, p = 0.40) and anecdotal for faces (346.3±9.29 vs. 347.9±9.23 ms; BF_{10} = 0.39; t(27) = -0.73, p = 0.23). Thus, RT data clearly indicate an attention capture by the memorized, but not by merely seen stimuli.

Dot-probe task - electrophysiological results

To further confirm the attention capture by stimuli maintained in WM, we analyzed the N2 posterior-contralateral (N2pc) ERP component, which is considered to be a robust index of covert attention shifts (Eimer, 1996; Kiss et al., 2008). To calculate N2pc we used signals from the P7/P8 electrodes in the 200 - 400 ms time window (**Fig. 3**), which is largely consistent with other studies using N2pc as measure of attention capture (e.g, Woodman et al., 2009; Reutter et al., 2017, Wójcik et al., 2019).

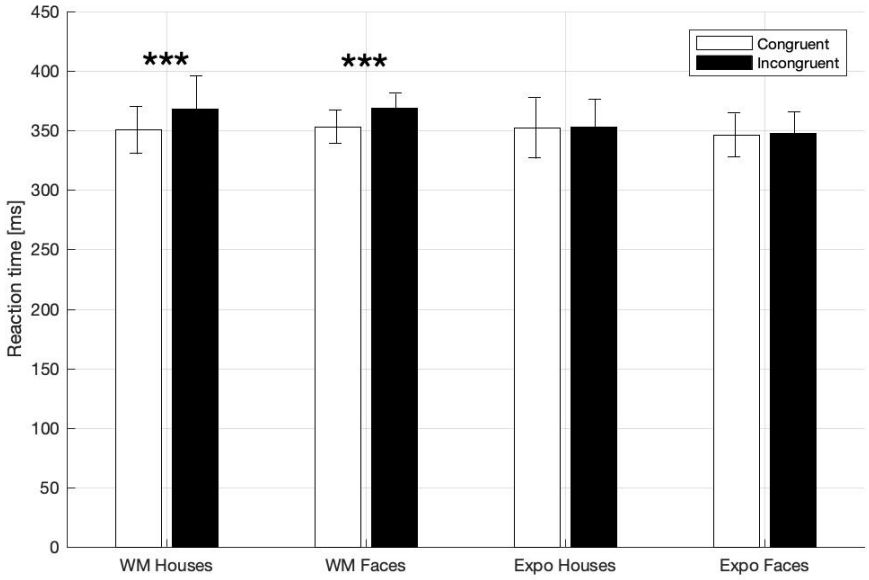

Figure 2. Mean RTs in dot-probe task for the WM and exposure conditions for both types of stimuli (House, Face). Asterisks indicate statistically significant difference between congruent and incongruent trials ($p < 0.001$). Error bars represent 2SEM.

We found anecdotal evidence that N2pc occurred in response to memorized houses (-0.14±0.08; BF_{10} = 1.32; t(27) = -1.69, p = 0.052), and strong evidence that it was evoked by memorized faces (-0.29±0.09; BF_{10} = 23.26; t(27) = 3.22, p = 0.002). However, in the mere exposure condition anecdotal evidence in favour of the null hypothesis (i.e. absence of N2pc) was observed for houses (-0.13±0.12; BF_{10} = 0.63; W= 123.00, p = 0.035), and moderate evidence in favour of the null hypothesis for faces (-0.04±0.09; BF_{10} = 0.31; t(27)= - 0.51, p = 0.306). Thus, in addition to behavioral results, electrophysiological data also indicate that stimuli actively maintained in working memory automatically capture attention, while a mere exposure of the same stimuli does not result in an attention capture. However, the effect is conclusive in case of faces, but anecdotal in case of houses.

We further conducted an additional exploratory analysis in the 100-200 ms time window, which was motivated by a post-hoc visual inspection of the obtained ERP waveforms. We found strong evidence that such an early contralateral negativity occurs for the memorized faces (-0.23±0.07; BF_{10} = 10.88; t(27) = -3.18, p = 0.004). However, such an early effect was not observed in other conditions, neither for memorized houses (0.01±0.08; BF_{10} = 0.20; t(27) = 0.06, p = 0.95), nor for merely seen faces (0.06±0.06; BF_{10} = 0.32; t(27) = 1.01, p = 0.32) or houses (0.00±0.08; BF_{10} =0.20; W = 201, p = 0.97), with Bayes factor indicating moderate evidence in favor of the null hypothesis in all three cases. Thus, the faces actively maintained in WM not only cause a robust attention capture indicated by the N2pc component, but they also evoke a lateralized activity in a very early time window.

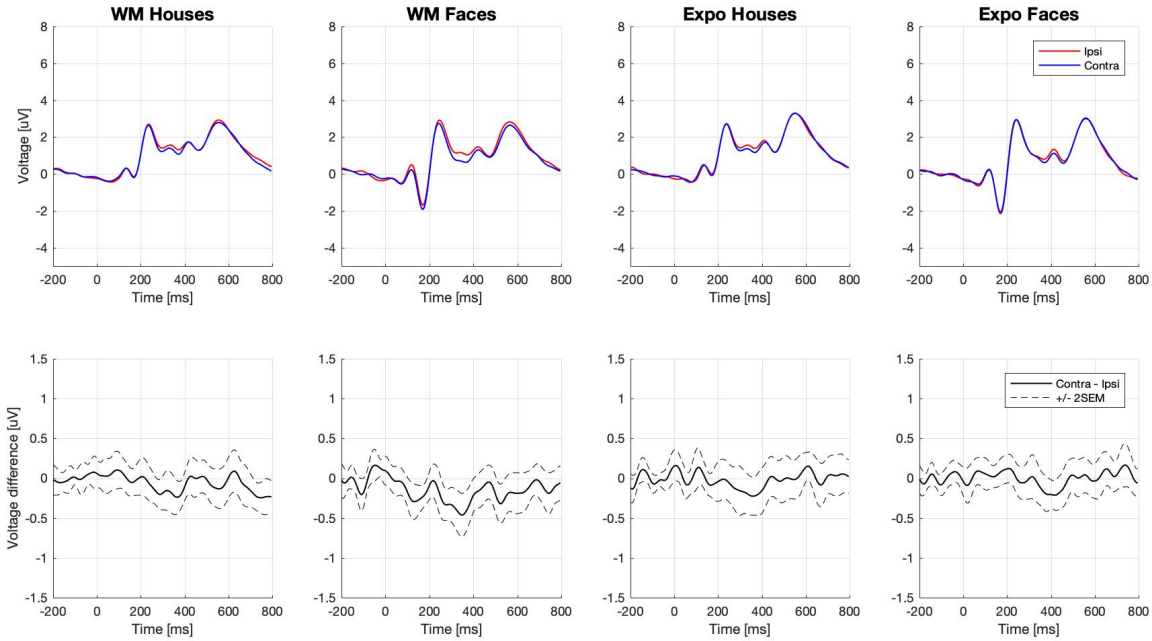

Figure 3. Event related potentials in the dot-probe task. Electrodes P7/P8 were chosen for the analysis. Waveforms from electrodes ipsi- and contra-lateral to the seen and memorized stimuli are presented in the top row. Difference waveforms (i.e. contra - ipsi-lateral) are presented in the bottom-row.

Discussion

The present study examined whether complex naturalistic stimuli that are actively maintained in WM attract visual attention upon subsequent presentation. Such a WM-guided attentional selection has been consistently observed for simple stimuli (review: Soto et al., 2008), but it was not known whether a similar mechanism can operate with complex stimuli as memory items. To address this question we asked subjects to memorize images of faces and houses, which were then presented as distractors in the dot-probe task. Our main finding is that in the dot-probe task RTs were significantly shorter when the target dot followed a memorized face or house, in comparison to the situation when it followed a control stimulus. This indicates that attention was indeed captured by the memorized stimuli. Importantly, the RT effect was not observed, neither for faces nor for houses, when subjects merely saw the template images without an instruction to memorize them. Thus, our results are in line with the visual search experiments, which showed increased RTs when a simple stimulus maintained in WM appeared in the search array as a distractor (Soto et al., 2005, 2007, 2008, 2009; Olivers et al., 2006). Further, the WM-based attention capture observed in our study was automatic and involuntary, as it occurred despite being detrimental to the task performance (again, in agreement with previous

studies using simple items; Olivers et al., 2011). Therefore, our study replicates and extends the scope of the WM-based attention guidance effect and, in more general terms, provides further evidence that complex stimuli are able to guide attention in an automatic manner (see: Wolfe and Horowitz, 2017).

Importantly, several previous studies reported that complex but artificial shapes (Downing and Dodds, 2004; Zhang et al., 2010) or drawings of real life objects (Houtkamp and Roelfsema, 2006) did not attract attention when maintained in WM. This led Zhang et al. (2010) to conclude that the attentional guidance critically depends on the stimulus features, with a stronger effect for simple than for complex ones. However, both faces and houses are defined by multiple features and by relations among them, and thus can be considered even more complex than the stimuli used previously. In our study both categories attracted attention, which indicates that the mere complexity of a stimulus is not likely to be a critical factor and, consequently, challenges the conclusion of Zhang and colleagues (2010). We rather argue that, in contrast to the artificial ones our stimuli might have been processed holistically and perceived as more meaningful. This is in line with the view of Xu (2017), arguing that visual WM is not typically used to encode features of a single dimension, but rather to store integrated representations of meaningful objects. Thus, our results obtained with face and house images reveal that involuntary influences of WM on attention might play a role and be further investigated also in more ecological conditions.

Two opposing accounts of the mechanism behind the WM-based guidance effect have been proposed. First one emphasizes the role of verbal (and perhaps semantic) representations in WM maintenance and subsequent directing of visual attention. It is based on studies showing that verbalization by itself can induce attention guidance (Soto and Humphreys, 2007), and that even when visual stimuli are used as memory items the articular suppression task impairs the guidance effect (Soto and Humphreys, 2008; Downing and Dodds, 2004; Woodman and Luck, 2007). However, the second view assumes that the guidance effect relies predominantly on visual representations. It is supported by experiments revealing the effect only when stimuli were defined by small and hard to verbalize differences in their attributes (e.g. hues of one color or slightly differing shapes) but not when easy to verbalize categorical differences were used (Olivers et al., 2006). Importantly, the stimuli used in our study were also difficult to verbalize and required maintaining a predominantly visual representation. Thus, our results provide further support for the second view.

Collected electrophysiological data allowed us to evaluate the time-course of attentional selection. ERP analysis revealed presence of the N2pc component - a posterior negativity contralateral to the memorized image occuring in the 200-400 ms temporal window - indicative of attentional prioritization of the memorized images (Eimer, 1996; Kiss et al., 2008). However, while the RT analysis indicates a robust attention capture for both stimuli types, Bayesian analysis of the N2pc component provides conclusive evidence in case of faces, but not houses. Speculatively, the following reason might account for the discrepancy between RT and ERP data in the houses WM condition: attention was shifted robustly to the memorized houses and consequently caused the RT effect, but the

exact timing of the stimulus recognition and attention shift might have varied from trial to trial, resulting in a poor signal-to-noise ratio in the averaged ERP (while recognition of faces might have been more automatic and more precisely time-locked to the stimulus onset). Relatedly, whereas in the present study attention capture was observed for task-irrelevant stimuli (and is therefore interpreted as automatic), recent studies suggest that both automatic and strategic processes mediate the WM-based guidance effect, and that it is stronger when stimuli are task-relevant (Carlisle and Woodman, 2011, 2013). Therefore, investigating the effect of naturalistic stimuli in situations when they are task-relevant, which would also more closely reflect daily life situations, might reveal even stronger effects.

Faces, due to their evolutionary and social importance, are processed in a largely automatic and holistic manner (Farah et al., 1998; Tsao and Livingstone, 2008). Importantly, faces benefit from a WM advantage, with holistic encoding strategies thought to allow overcoming the limited capacity of the WM system (Curby et al., 2009). In our study we found only anecdotal evidence that faces were memorized better than houses, but any conclusions must be treated with caution as the stimuli were not appropriately matched for such direct comparisons between categories. Further, recent continuous flash suppression (CSF) studies show that faces actively maintained in WM break the CFS faster than faces that were merely seen (Gayet et al., 2013; Pan et al., 2014). The fact that in the CFS paradigm WM can bias face perception outside of awareness is in line with the automatic and involuntary (possibly pre-attentive) nature of the effect found here. Further, the attention capture revealed in the present study is a plausible mechanism accounting for the CFS effects.

Finally, face recognition is performed by a specialized set of brain regions (Haxby et al., 2000; Kanwisher and Yovel, 2006) with the initial stages of face categorization occurring as early as 80-150 ms post-stimulus (Herrmann et al., 2005). This is in line with our exploratory analysis, in which we found evidence for a very early prioritization of the memorized faces. While the N2pc component typically emerges around 200 ms after the stimulus onset (e.g. Woodman et al., 2009; Reutter et al., 2017, Wójcik et al., 2019), the memorized faces caused a contralateral negativity already around 100 ms (i.e. co-occurring with the P100 component). While it remains to be investigated what processes underlie such an early effect, we consider our finding relevant to the ongoing debate on the neuronal mechanisms of visual working memory. Here, two opposing theories have been proposed: first, the top-down amplification hypothesis, which assumes that visual WM items are maintained by fronto-parietal interactions (Bettencourt and Xu, 2016; Christophel et al., 2018, Riley and Constantinidis, 2016, Thigpen et al., 2019; review: Xu et al., 2017); second, the sensory-recruitment hypothesis, assuming that visual WM items are stored and maintained in the visual cortex (i.e. that perception and visual WM share the same neural substrate; Postle, 2006). The latter view might particularly effectively account for the automatic interactions between perception and the WM contents, which were observed in our and other studies (e.g., Albers et al., 2013, Gayet et al., 2017, Silvanto and Cattaneo, 2010, Teng and Kravitz, 2019). Importantly, given that ERP components observed in the

100-200 ms time-range are generated by sensory brain regions and reflect perceptual processing (Nusslock, 2016), such an early prioritization of the memorized faces provides support for the sensory recruitment theory. Further, such an early effect was not found in previous dot-probe studies using very salient and relevant emotional faces (Holmes et al., 2009) or self-faces of participants (Wójcik et al., 2019), which further indicates it might specifically reflect a match between the WM-maintained representation and an incoming stimulus. However, because this analysis was exploratory, the conclusion should be treated with caution.

In conclusion, by showing that an automatic guidance of attention from WM can operate on complex naturalistic stimuli our study further replicates and, at the same time, extends the investigated effect. Our results are thus consonant with emerging evidence for the prospective and dynamic role of memory in guiding perception and behaviour. Further, we provide evidence for a very early prioritization of the memorized face images, which resonates well with previous studies indicating preferential processing of faces. Finally, the present study encourages further investigation of the WM-based influences on attention in more ecological settings.

Footnotes

1. Female neutral face KDEF photographs were: AF01NES, AF02NES, AF03NES, AF04NES, AF05NES, AF06NES, AF07NES, AF08NES, AF09NES, AF10NES, AF11NES, AF12NES, AF13NES, AF14NES, AF16NES, AF17NES, AF18NES, AF19NES, AF20NES, AF21NES, AF22NES, AF23NES, AF24NES, AF25NES, AF26NES. AF27NES, AF28NES, AF29NES, AF30NES, AF31NES. Male neutral face KDEF photographs were: AM01NES, AM02NES, AM03NES, AM04NES, AM05NES, AM06NES, AM07NES, AM08NES, AM09NES, AM10NES, AM11NES, AM13NES, AM14NES, AM15NES, AM17NES, AM18NES, AM21NES, M22NES, AM23NES, AM25NES, AM26NES, AM27NES, AM28NES AM29NES, AM30NES, AM31NES, AM32NES, AM33NES, AM34NES, AM35NES

2. The Dalhouses photographs were: 1, 2, 3, 5, 6, 7, 12, 13, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 44, 45, 46, 47, 48, 49, 52, 61, 62, 64, 67, 68, 69, 71, 72, 73, 74, 75, 77, 78, 79, 81, 84, 85, 86, 87, 88

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427-1431.

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature neuroscience*, *19*(1), 150.

Carlisle, N. B., & Woodman, G. F. (2011). Automatic and strategic effects in the guidance of attention by working memory representations. *Acta psychologica*, *137*(2), 217-225.

Carlisle, N. B., & Woodman, G. F. (2013). Reconciling conflicting electrophysiological findings on the guidance of attention by working memory. *Attention, Perception, & Psychophysics*, *75*(7), 1330-1335.

Christophel, T. B., Allefeld, C., Endisch, C., & Haynes, J. D. (2018). View-independent working memory representations of artificial shapes in prefrontal and posterior regions of the human brain. *Cerebral Cortex*, *28*(6), 2146-2161.

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 94.

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

Downing, P., & Dodds, C. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*(6), 689-703.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and clinical neurophysiology*, *99*(3), 225-234.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is" special" about face perception?. *Psychological review*, *105*(3), 482.

Filliter, J. H., Glover, J. M., McMullen, P. A., Salmon, J. P., & Johnson, S. A. (2016). The DalHouses: 100 new photographs of houses with ratings of typicality, familiarity, and degree of similarity to faces. Behavior research methods, 48(1), 178-183.

Gayet, S., Paffen, C. L., & Van der Stigchel, S. (2013). Information matching the content of visual working memory is prioritized for conscious access. *Psychological Science*, *24*(12), 2472-2480.

Gayet, S., Guggenmos, M., Christophel, T. B., Haynes, J. D., Paffen, C. L., Van der Stigchel, S., & Sterzer, P. (2017). Visual working memory enhances the neural response to matching visual input. *Journal of Neuroscience*, *37*(28), 6638-6647.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223-233.

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates rapid eye movements to simple onset targets. *Psychological science*, *24*(5), 790-796.

Holmes, A., Bradley, B. P., KRAGH Nielsen, M., & Mogg, K. (2009). Attentional selectivity for emotional faces: Evidence from human electrophysiology. *Psychophysiology*, *46*(1), 62-68.

Herrmann, M. J., Ehlis, A. C., Ellgring, H., & Fallgatter, A. J. (2005). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *Journal of neural transmission*, *112*(8), 1073-1081.

Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 423.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109-2128.

Kiss, M., Jolicœur, P., Dell'Acqua, R., & Eimer, M. (2008). Attentional capture by visual singletons is mediated by top□down task set: New evidence from the N2pc component. *Psychophysiology*, *45*(6), 1013-1024.

Kumar, S., Soto, D., & Humphreys, G. W. (2009). Electrophysiological evidence for attentional guidance by the contents of working memory. *European Journal of Neuroscience*, *30*(2), 307-317.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of abnormal psychology*, *95*(1), 15.

Nobre, A. C., & Stokes, M. G. (2019). Premembering experience: a hierarchy of time-scales for proactive attention. *Neuron*, *104*(1), 132-146.

Nusslock, R. (2016). Neurophysiology and neuroimaging. In: J.C. Norcross, G.R. VandenBos, D.K. Freedheim, APA Handbook of Clinical Psychology: Vol. 2. Theory and Research. Washington DC: *American Psychological Association*.

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243.

Olivers, C. N., 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*(7), 327-334.

Pan, Y., Lin, B., Zhao, Y., & Soto, D. (2014). Working memory biasing of visual perception without awareness. *Attention, Perception, & Psychophysics*, *76*(7), 2051-2062.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23-38.

Reutter, M., Hewig, J., Wieser, M. J., & Osinsky, R. (2017). The N2pc component reliably captures attentional bias in social anxiety. *Psychophysiology*, *54*(4), 519-527.

Riley, M. R., & Constantinidis, C. (2016). Role of prefrontal persistent activity in working memory. *Frontiers in systems neuroscience*, *9*, 181.

Schneegans, S., Spencer, J. P., Schöner, G., Hwang, S., & Hollingworth, A. (2014). Dynamic interactions between visual working memory and saccade target selection. *Journal of vision*, *14*(11), 9-9.

Silvanto, J., & Cattaneo, Z. (2010). Transcranial magnetic stimulation reveals the content of visual short-term memory in the visual cortex. *Neuroimage*, *50*(4), 1683-1689.

Silvis, J. D., & Van der Stigchel, S. (2014). How memory mechanisms are a key component in the guidance of our eye movements: Evidence from the global effect. *Psychonomic bulletin & review*, *21*(2), 357-362.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in cognitive sciences*, *12*(9), 342-348.

Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 730.

Soto, D., & Humphreys, G. W. (2008). Stressing the mind: The effect of cognitive load and articulatory suppression on attentional guidance from working memory. *Perception & Psychophysics*, *70*(5), 924-934.

Soto, D., & Humphreys, G. W. (2009). Automatic selection of irrelevant object features through working memory: Evidence for top-down attentional capture. *Experimental psychology*, *56*(3), 165-172.

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature human behaviour*, *3*(8), 827-836.

Thigpen, N., Petro, N. M., Oschwald, J., Oberauer, K., & Keil, A. (2019). Selection of visual objects in perception and working memory one at a time. *Psychological science*, *30*(9), 1259-1272.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annu. Rev. Neurosci.*, *31*, 411-43

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search?. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 363.

Woodman, G. F., Arita, J. T., & Luck, S. J. (2009). A cuing study of the N2pc component: An index of attentional deployment to objects rather than spatial locations. *Brain research*, *1297*, 101-111.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1-8.

Wójcik, M. J., Nowicka, M. M., Bola, M., & Nowicka, A. (2019). Unconscious detection of one's own image. *Psychological science*, *30*(4), 471-480.

Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends in Cognitive Sciences*, *21*(10), 794-815.

Zhang, B., Zhang, J. X., Kong, L., Huang, S., Yue, Z., & Wang, S. (2010). Guidance of visual attention from working memory contents depends on stimulus attributes. *Neuroscience Letters*, *486*(3), 202-206.