CORTEX 122 (2020) 140–158

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

Emotional capture during emotion-induced blindness is not automatic

James E. Hoffman [a,*], Minwoo Kim [a], Matt Taylor [a] and Kelsey Holiday [b]

[a] University of Delaware, Newark, DE, USA
[b] San Diego State University, CA, USA

ARTICLE INFO

Article history:
Received 29 June 2018
Reviewed 27 September 2018
Revised 1 February 2019
Accepted 15 March 2019
Published online 27 March 2019

Keywords:
Emotion induced blindness
Emotional attention capture
Multiple object tracking
Automaticity
Early posterior negativity

ABSTRACT

The present research used behavioral and event-related brain potentials (ERP) measures to determine whether emotional capture is automatic in the emotion-induced blindness (EIB) paradigm. The first experiment varied the priority of performing two concurrent tasks: identifying a negative or neutral picture appearing in a rapid serial visual presentation (RSVP) stream of pictures and multiple object tracking (MOT). Results showed that increased attention to the MOT task resulted in decreased accuracy for identifying both negative and neutral target pictures accompanied by decreases in the amplitude of the P3b component. In contrast, the early posterior negativity (EPN) component elicited by negative pictures was unaffected by variations in attention. Similarly, there was a decrement in MOT performance for dual-task versus single task conditions but no effect of picture type (negative *vs* neutral) on MOT accuracy which isn't consistent with automatic emotional capture of attention. However, the MOT task might simply be insensitive to brief interruptions of attention. The second experiment used a more sensitive reaction time (RT) measure to examine this possibility. Results showed that RT to discriminate a gap appearing in a tracked object was delayed by the simultaneous appearance of to-be-ignored distractor pictures even though MOT performance was once again unaffected by the distractor. Importantly, the RT delay was the same for both negative and neutral distractors suggesting that capture was driven by *physical salience* rather than *emotional salience* of the distractors. Despite this lack of emotional capture, the EPN component, which is thought to reflect emotional capture, was still present. We suggest that the EPN doesn't reflect capture but rather downstream effects of attention, including object recognition. These results show that capture by emotional pictures in EIB can be suppressed when attention is engaged in another difficult task. The results have important implications for understanding capture effects in EIB.

Published by Elsevier Ltd.

* *Corresponding author*. Department of Psychology, University of Delaware, Newark, DE, 19716-2577, USA.
E-mail address: hoffman@psych.udel.edu (J.E. Hoffman).
https://doi.org/10.1016/j.cortex.2019.03.013
0010-9452/Published by Elsevier Ltd.

1. Introduction

Visual scenes are rich sources of information and our ability to comprehend them and respond appropriately to the objects they contain depends on attentional mechanisms that prioritize some objects over others. Visual attention mechanisms are typically categorized as being top-down or voluntary versus bottom-up or automatic (Theeuwes, 2010). Looking for one's car in a crowded parking lot is an example of top-down attention. Having one's attention captured by a sudden movement in an otherwise static display illustrates bottom-up or automatic attention. According to Theeuwes (2010), salient objects capture attention automatically.

In the laboratory, several paradigms have been developed to study automatic attention capture. For example, in the additional singleton paradigm (Theeuwes, 1992), observers might search for a green, diamond shaped target in a display of green circles. In this case, the target is a said to be a "shape singleton" because it is the only object with a different value on the shape dimension. On some trials, one of the green circles is replaced by a red circle. Now there are two singletons in the display (a target and a distractor) and the red distractor will be more salient than the target because color singletons tend to have greater salience than shape singletons. Even though participants know they are supposed to ignore the red distractor, they are unable to do so and it captures attention resulting in a slower response to the target compared to trials without the singleton distractor. According to the *stimulus-driven capture hypothesis* (Theeuwes, 2010), spatial attention is automatically allocated to the location of the most salient object in the display with salience being determined by the physical features in the display. After attention is allocated to the salient distractor, additional processing may reveal that is isn't the target, in which case, attention will be disengaged and then allocated to the next most salient object in the display which will usually be the target. Theeuwes (2010) suggests that when observers are experienced with this task, the time that attention dwells on the distractor can be relatively short because people can quickly determine that the distractor is not the target. These short dwell times are what underlie the relatively small capture effects in which the time to respond to the target is delayed by about 25 msec in the presence of a distractor.

Other studies however have shown that capture by salient singletons can be modulated by top-down attention. For example, Folk, Remington, and Johnston (1992) showed that if observers adopt a perceptual set for a particular feature such as color, they are not distracted by distractors that are shape or motion singletons. They called this the *contingent capture hypothesis* which holds that features only attract our attention if they match the perceptual set we have adopted. Notice that the contingent capture hypothesis is diametrically opposed to the stimulus driven capture hypothesis and they make very different testable claims about what one should observe in capture experiments. Nonetheless, several decades of studies have failed to convincingly support one hypothesis over the other. Recently, however, Luck and colleagues (reviewed in Gaspelin, Leonard, & Luck, 2017) have introduced a new hypothesis of *signal suppression* that appears to largely resolve this debate. Some of their evidence rests on studies using event related brain potentials (ERPs) to which we turn next.

The N2pc component (Luck, 2012) has been instrumental in uncovering attentional mechanisms underlying capture. It is a negative component with a peak latency of approximately 200–300 msec. appearing over posterior, contralateral (the "pc" in N2pc) sensors. Extensive research supports the claim that the N2pc is related to attentional selection of objects. For example, when observers are searching for a salient target object, an N2pc component will be elicited over the hemisphere that is contralateral to the visual field in which the target appears (Eimer, 2015). This component can also be observed in response to task-irrelevant, salient distractors in the additional singleton paradigm (Hickey, Di Lollo, & McDonald, 2009) reflecting the capture of spatial attention by the salient distractor. However, the N2pc may also be followed by a positive component with a similar scalp topography known as the P_D component (positivity to a distractor). The P_D reflects a suppression of the distractor object (Sawaki, Geng, & Luck, 2012) which would presumably make it easier to disengage attention from the distractor in order to allocate it to the target. In some cases, the P_D may occur early enough to suppress the distractor before it can capture attention, thereby eliminating the deleterious effects of the salient distractor on the speed of responding to the target (Sawaki & Luck, 2010). The signal suppression hypothesis is similar to the contingent capture claim that top–down processes can affect capture, but it is also consistent with the stimulus driven hypothesis which predicts that in the absence of suppression, salient objects will capture attention even if they do not match the current selection goals.

Awh, Belopolsky, and Theeuwes (2012) pointed out that the basic dichotomy between top-down and bottom-up mechanisms of attention allocation was inadequate to explain the full range of influences on attention allocation. They suggested that this dichotomy should be supplemented by effects of selection history and reward. For present purposes, reward effects are particularly important. For example, Moray (1959) reported that participants were often aware of the occurrence of their own names in an unattended channel during auditory shadowing. Apparently, because our own names are "important" they capture attention automatically. This capture can't be explained either by bottom-up salience (our own names don't possess basic auditory features that differ from other spoken words) or top-down attention (participants in the experiment didn't have a current goal of listening for their own name). Emotional stimuli, such as threatening faces, presumably capture attention for similar reasons. In this case, we might say that they have emotional salience rather than physical salience. There is, of course, good reason to attend to emotional stimuli even when they aren't relevant to our current task. Negative emotional stimuli, such as dangerous animals, an angry face, or a severely injured person signal the presence of threat and associated harm. Detecting the presence of threatening stimuli can allow the formulation of plans for escape, defense, etc. and such a system would have clear survival value that would confer an evolutionary advantage (Öhman, Flykt, & Esteves, 2001). According to Awh et al., these different forms of salience may all be represented in a single

priority map which represents the relative salience and locations of all the stimuli appearing at any given time. Allocation of attention would be guided to the most salient location in this map.

Although emotional stimuli and physically salient, unemotional stimuli may share a common priority map, it isn't clear how similar they are in terms of underlying mechanisms. In particular, can top-down mechanisms of attention prevent emotional capture as they can in the case of physical salience driven capture (reviewed above)? In the case of physical salience, the visual features responsible for capture are processed in "pre-attentive vision", during the initial feedforward sweep of activity in visual cortex (Theeuwes, 2010). In contrast, emotional salience depends on knowing more abstract features associated with a stimulus, such as is its emotional valence, which would depend on stored knowledge about the object. This might require an initial object recognition stage followed by retrieval of the object's "meaning" which would require neural systems "higher" in the visual processing hierarchy than those associated with feature processing. In this case, one might argue that emotional salience capture might be *more vulnerable* to top-down effects than physical salience capture.

However, rapid and automatic processing of the valence of objects might be accomplished by sub-cortical circuits involving the amygdala that are specialized for processing emotional stimuli (Méndez-Bértolo et al., 2016; Pourtois, Schettino, & Vuilleumier, 2013; Rafal et al., 2015). This circuit is thought to operate in parallel with the main visual pathway that proceeds through the lateral geniculate nucleus to the striate and extrastriate areas that ultimately leads to storage of information in working memory as well as awareness. The amygdala pathway might operate outside the limited-capacity bottleneck that characterizes processing in the geniculo-striate system to provide rapid processing of emotional stimuli even when participants are engaged in other attention-demanding activities.

Because the amygdala has extensive connections with several cortical areas, it would have the ability to direct attention to an emotional object or event, resulting in rapid interference with processing of task-relevant stimuli. Several aspects of this proposal, however, remain controversial. For example, Pessoa and Adolphs (2010) pointed out that neuroimaging data are mixed regarding the claim of particularly rapid processing in the amygdala. Single unit recordings in monkey have shown that activity discriminating between emotional and non-emotional stimuli can reach prefrontal areas of the brain with comparable latency to activity in the amygdala.

In addition, findings are mixed regarding the question of whether emotional stimuli invariably capture attention and interfere with processing of task-relevant stimuli. For example, Vuilleumier, Armony, Driver, and Dolan (2001) showed that to-be ignored angry faces slowed judgments of simultaneously presented task-relevant houses and this interference was accompanied by activity in the amygdala and fusiform gyrus. However, Pessoa, McKenna, Gutierrez, and Ungerleider (2002) suggested that the house judgment task was too easy and didn't require full attention, allowing unused attentional resources to be allocated to the faces, a claim that is consistent with perceptual load theory (Lavie, 1995). When they used a more difficult line orientation judgment as the attended task, interference by emotional pictures, as well as valence-specific brain activation, was eliminated.

Research on emotional capture has also identified several ERP components that may be signatures of the capture process. For example, emotional pictures elicit an N2 component over posterior areas of the scalp with a latency of approximately 200–300 msec. This component is larger for emotional pictures compared to neutral control pictures which is assumed to reflect greater capture of attention by negative pictures (Schupp, Flaisch, Stockburger, & Junghöfer, 2006; Schupp, Stockburger, Bublatzky et al., 2007; Schupp, Stockburger, Codispoti et al., 2007). The negative-neutral difference wave is a useful measure of this contrast and is known as the early posterior negativity or EPN. The relationship between the EPN and N2pc components is currently unclear. They have similar scalp topographies and latencies and they both seem to index attention. In addition, the EPN isn't tied to emotional stimuli because similar N2 components are elicited by task-irrelevant emotional stimuli and task-relevant unemotional stimuli (Kennedy, Rawding, Most, & Hoffman., 2014). However, it is isn't known whether the EPN is lateralized like the N2pc. Several previous studies have suggested that the EPN component elicited by emotional pictures is unaffected by variations in voluntary attention and therefore reflects the automatic nature of attention capture by emotional stimuli (e.g., Holmes, Nielsen, Tipper, & Green, 2009; Rellecke, Sommer, & Schacht, 2012) although other studies have reported a reduction in the EPN to emotional pictures when attention is fully engaged on another difficult task (Schupp, Stockburger, Bublatzky et al., 2007; Wiens & Syrjänen, 2013).

A second component often elicited in emotional capture paradigms is the P3b which is broadly distributed over posterior, central and temporal areas. Its latency varies widely depending on the task but typically lies in the range of 300–700 msec. Like the EPN, it is larger for negative compared to neutral pictures (Hajcak, Weinberg, MacNamara, & Foti, 2011; Kennedy, Rawding, Most, & Hoffman, 2014). The underlying cognitive process associated with the P3b is still debated but popular candidates are decision making, consolidation of information into working memory (Donchin & Coles, 1998) and conscious awareness (e.g., Sergent, Baillet, & Dehaene, 2005).

The current study examines the issue of the automaticity of emotional capture using a variant of the emotion-induced blindness paradigm (EIB; Most, Chun, Widders, & Zald, 2005) which produces particularly large capture effects. In the typical EIB study, participants search a rapidly presented stream of upright scene pictures for a target picture (a scene picture that has been rotated to the right or left). The target picture is sometimes preceded by a task-irrelevant emotional or neutral picture. When the target is preceded by a *task-irrelevant* emotional distractor picture, target discrimination is impaired by as much as 40% while the neutral distractor produces a much smaller impairment of approximately 10%. The claim of automaticity is based on the fact that participants have a perceptual set to attend to a rotated scene picture (the target) but this top-down set isn't sufficient to prevent emotional

pictures from capturing attention and blocking access by the closely following target to later, limited-capacity mechanisms. Similar effects occur for positive, arousing emotional stimuli and these interference effects even occur in the face of high rewards (up to $90!) for ignoring them (Most, Smith, Cooter, Levy, & Zald, 2007). In addition, employing the same stimuli used in EIB studies, Choisdealbha, Piech, Fuller, and Zald (2017) reported that emotional pictures that appeared as second targets in an attentional blink (AB) paradigm, evaded the blink and in fact, interfered with perception of the preceding target. In a recent review of research on EIB, McHugo Olatunji and Zald (2013) suggest that EIB "provides a unique and robust measure of attentional capture", and they note that "at least to date, there is little evidence that the EAB (note: same as EIB) can be overcome by the application of top-down control".

The small blink effect for neutral pictures in EIB reflects a potentially important but often overlooked aspect of the stimuli used in EIB studies. Both negative and neutral distractors consist of close-ups of people and animals while the background pictures are wide-angle views of landscapes and cityscapes. This means that negative and neutral pictures are *relatively similar to each other* and *both are physically different than the background pictures* raising the issue of whether the observed capture effects in EIB can be attributed to physical salience. However, the fact that neutral pictures produce a smaller blink than emotional pictures is used as evidence that even if neutral pictures produce a small blink based on physical salience, emotional pictures capture *more attention* than neutral pictures, resulting in a larger blink. This appears to be consistent with the observation that emotional pictures also produce larger N2 components than neutral pictures (the EPN component) which supports the claim of greater attention capture. Greater *initial capture* of attention based on emotional valence would support the claim that emotional salience, like physical salience, is processed pre-attentively and can automatically capture visual attention.

An alternative view of the role of physical salience-driven capture in EIB is that initial capture of attention by negative and neutral pictures is the same and is based on physical salience. The reason that negative pictures produce larger EIB effects would be due to later processes, located downstream from the attentional capture stage, which are engaged by emotional but not neutral stimuli. This claim is consistent with a study by Kennedy et al. (2014) who examined ERPs in a standard EIB paradigm. They found that emotional pictures produced a P3b component while neutral pictures did not. In addition, negative pictures produced more interference when they elicited a P3b, indicating that suppression of the target in EIB is related to negative pictures occupying late, limited-capacity systems. The same is true of the attentional blink (AB): the second target is suppressed on trials where the first target elicited a P3b (Kranczioch, Debener, Maye, & Engel, 2007; Shapiro, Schmitz, Martens, Hommel, & Schnitzler, 2006). These results suggest that a major source of interference in EIB *and* AB is the ability of an initial stimulus (task-irrelevant emotional pictures in the case of EIB and a task-relevant unemotional picture in AB) to occupy late stages involved with working memory consolidation and awareness that prevents access by shortly following targets. In contrast, Kennedy at al. found the amplitude of the N2 elicited by negative pictures was approximately the same for correct and incorrect trials, although it was somewhat prolonged on error trials. Similarly, MacLeod, Stewart, Newman, and Arnell (2017) failed to find any relationship between the amplitude of the N2 elicited by emotional words and suppression of the following target word. In summary, there is currently no strong evidence for the popular assumption that emotional pictures *initially capture* more attention than neutral pictures in EIB leaving open the possibility that initial, attentional capture in this paradigm is the same for negative and neutral distractors and is driven by physical salience rather than emotional salience. The larger blink associated with negative pictures may be due to their greater engagement of *late processes* that are located downstream from the early visual areas responsible for computing physical salience.

This paper examines the issue of whether attentional capture in the EIB paradigm is completely automatic in the sense that it occurs even when emotional pictures are task-irrelevant, and participants are engaged in a difficult primary task at the moment the pictures are presented. In addition to behavioral measures of capture, we examined ERPs to help determine whether early or late processing stages are affected by top-down attention. In the first experiment, participants identified negative and neutral pictures in a RSVP stream while tracking multiple objects moving in front of the pictures (see Cohen, Alvarez, & Nakayama, 2011; D'Andrea-Penna, Frank, Heatherton, & Tse, 2017). We examined the effect of changing task priority on the ability of emotional pictures to capture attention. In one condition, participants had to identify a negative or neutral picture appearing in a rapid serial visual presentation (RSVP) stream of background scene pictures. This condition should reveal ERP components elicited by emotional pictures when they are fully attended and task-relevant. In a second condition, participants were instructed to ignore the picture stream and attempt to track three target objects moving among three identical distractors (MOT). This condition should reveal which ERP components are affected by withdrawal of attention from the negative picture. Finally, in a dual task condition, participants were instructed to treat the MOT task as primary and to identify target pictures in the stream as long as it didn't interfere with the primary task. This condition should reveal if even minimal attention might be sufficient for full processing of emotional pictures. If negative pictures automatically capture attention, as suggested by EIB studies, we should find that negative pictures impair performance on the MOT task more than neutral pictures even when participants are instructed to allocate all of their attention to the MOT task.

2. Experiment 1

2.1. Methods

No part of the study procedures or analyses for Experiments 1 and 2 were preregistered prior to research being undertaken. Sample size was determined based on previous studies in our lab using similar procedures and measures. Inclusion/Exclusion criteria were determined prior to data collection.

2.1.1. Participants

Twenty-two right-handed, neurologically normal participants (15 women, 7 men, mean age: 20.8, age range: 19–28) were recruited through a university-sponsored classified ad. Two participants were excluded due to experimenter error and another two were excluded for excessive artifacts in the ERP recording (rejection of more than 30% of total epochs), resulting in a total of 18 participants. The University of Delaware Institutional Review Board approved the study and all participants provided informed consent. They all reported normal or corrected-to-normal vision and were remunerated at the rate of $10 per hour for their participation.

2.1.2. Apparatus

Observers were tested in a dimly lit and acoustically isolated room. Displays were presented on a SAMSUNG 2233RZ 22" LCD Monitor (Wang & Nikolic, 2011) having 1,680 × 1,050 pixel resolution at a 120 Hz refresh rate. Viewing distance was fixed at 74 cm using a chin rest and the view angle was horizontally matched to the center of the monitor. The experiment was controlled by a Dell 3.60 GHz computer and programmed using Blitz 3D software (Sibly, 2005). Eye blinks and movements were monitored with an Eyelink 1000 eye tracker (SR Research, Ontario, Canada), which sampled eye position at 500 Hz. An eye movement was defined as three consecutive eye samples that were more than 1.4 degrees of visual angle from fixation. The electroencephalogram (EEG) was recorded with an Electrical Geodesics Inc. system Net Amps 200 using 128 channel Hydrocel Sensor Nets and EGI Net Station 4.5.6 software.

2.1.3. Stimuli

The stimuli appeared on a black screen containing a .37° × .37° fixation-cross which remained visible for the duration of the trial. Displays consisted of a stream of pictures centered on the fixation point. Each picture was 9.6° × 6.4° in size and shown for 100 msec, followed immediately by the next picture in the sequence. The background pictures were drawn randomly from a set of 252 landscape and architectural photographs (see Fig. 1). In two thirds of the sequences, a picture containing a person or animal replaced one of the background pictures. We refer to these as "animate pictures" and they could have either negative or neutral valence. There were 43 negative pictures containing depictions of violence, medical trauma, or threatening animals and 52 neutral pictures that showed people and animals in a neutral context. The majority of the animate pictures were taken from the International Affective Picture System (IAPS), a set of picture stimuli normalized and rated for arousal and emotional valence (Lang, Bradley, & Cuthbert, 2008). This set was supplemented with 16 additional pictures chosen from the internet. Streams without an animate picture contained only baseline landscape or architectural pictures and were used as a control condition.

We collected affective ratings of our picture set from 36 additional participants using an online testing service (Mturk) as well as studies in our lab using the departmental subject pool. Data from 11 subjects who failed to follow instructions, were excluded from the analysis, leaving a total of 25 participants (14 men, mean age: 25.76). Participants from Mturk were compensated at a rate of $6 while volunteers from the subject pool received two hours of participation credit. Participants used on-screen, moving sliders to rate the valence (1: positive, 5: neutral, 9: negative) and arousal (1: low arousal, 5: neutral, 9: high arousal) of the pictures which remained on the screen until both responses were submitted. Dependent measures t-tests revealed that the average valence of the negative pictures ($M = 7.92$, $SE = .15$) was significantly more

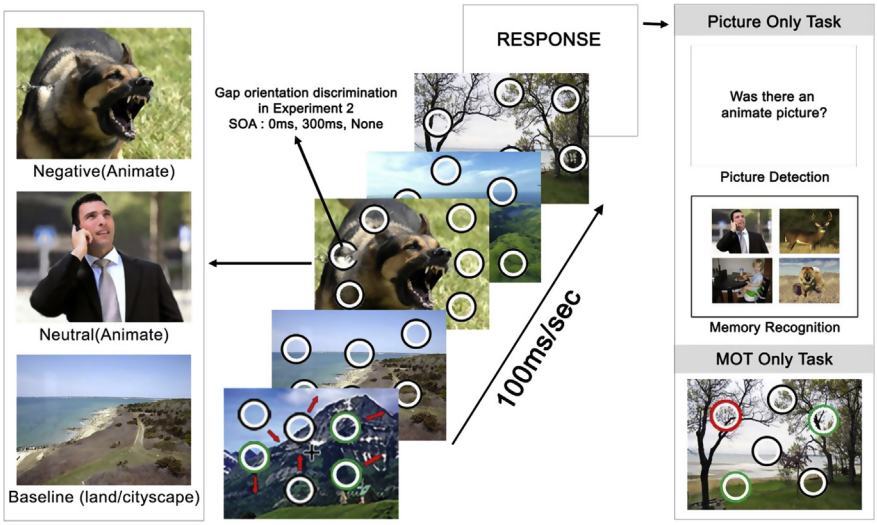

Fig. 1 – An illustration of a typical trial in Experiment 1. Six disks moved in front of an RSVP stream of background pictures portraying landscapes and cityscapes. The stream also contained one of three kinds of distractors: a negative or neutrally valenced picture containing people or animals (animates), or a baseline picture drawn from the same category as the background pictures. There were three attention tasks. In the Picture-Only condition, participants had to detect and recognize the animate pictures. In the MOT-Only condition, they tracked three disks (shown in green) while ignoring the pictures. In the Dual-Task condition, they performed both tasks while emphasizing MOT.

negative, $t(93) = 19.91$, $p < .001$, than the neutrals ($M = 4.83$, $SE = .06$). Negative pictures ($M = 6.26$, $SE = .49$) also had significantly higher arousal, $t(93) = 26.04$, $p < .001$, compared to neutrals ($M = 4.04$, $SE = .07$).

In addition to the pictures, displays contained a set of six identical disks (1° in diameter), each consisting of two concentric rings, one white and one black, which made them visible on both light and dark areas of the pictures. The disks appeared in front of the pictures and moved along independent, random trajectories with a constant velocity of 7.2°/s. They bounced off each other as well as the sides of the picture frame (see Fig. 1).

2.1.4. *Procedure*

Subjects were shown a sample of the negative pictures prior to the experiment and were informed they could terminate participation at any point. At the beginning of the experiment, each subject completed 15 practice trials (5 per task condition) with an RSVP rate of 5 images per second under supervision of the experimenter. The RSVP rate increased to 10 images per second for the actual experiment. Each trial began with a centrally presented red fixation cross. Participants were instructed to maintain fixation on the cross at the center of the screen for the duration of the trial and refrain from blinking or making unnecessary movements. Participants initiated the trial by pressing the left mouse button which displayed six stationary disks and initiated the RSVP sequence of 20 images. After 500 msec, the disks began moving along random trajectories. A single picture containing an animate (person or animal, negative or neutral) could appear randomly in positions 6–12 of the RSVP sequence. Participants were provided feedback about eye movements or blinks at the end of the trial.

In the Picture-Only task, subjects were told to ignore the moving disks and detect the animate picture in a stream of "scene pictures" (wide-angle views of landscapes and cityscapes). At the end of each trial, the RSVP stream and the disks disappeared, and subjects were prompted to indicate whether the sequence of images included an animate picture by pressing a mouse button (left for "yes" and right for "no"). A maximum of one animate picture was present in each sequence. "Correct" or "Incorrect" feedback was provided immediately after the response. If an animate image was present, participants were then asked to use the mouse to select the matching picture from of a set of 4 pictures, regardless of the preceding detection response. The choices were randomly drawn from the same set of pictures as the presented picture (i.e., neutral or negative) with the location of the matching picture randomized across the set of alternatives. A colored box appeared around the image to provide the subject with feedback on their response (green for correct and red for incorrect). At the completion of the trial, the screen automatically reverted to its initial state. One third of the sequences did not contain an animate picture and are referred to as the Baseline condition.

In the MOT-Only task, subjects were instructed to attend only to the moving circles and ignore the picture sequence. When subjects clicked the mouse to start each trial, six stationary disks appeared with the three disks to be tracked highlighted in green. After 500 msec, all disks reverted to black and white and began to move around the screen. At the end of the sequence, participants indicated the target circles by clicking on them using the mouse. Each chosen disk turned green if the response was correct and red if it was incorrect. A total of three responses was required on each trial.

In the Dual-Task condition, subjects were asked to perform both tasks on each trial with emphasis placed on accurate tracking of the circles as the primary task. At the end of the trial, picture detection and recognition responses followed the choice of MOT targets.

The experiment consisted of 540 trials divided into nine blocks consisting of three repetitions of the Picture-Only, MOT-Only and Dual-Task conditions in that order. Short breaks were given between trials and after each block.

2.1.5. *Electrophysiological recording and analysis*

As recommended by the manufacturer, individual electrode impedances were kept below 75 kΩ. The data were referenced online to the vertex, band-pass filtered (.01–80 Hz), and digitized at 200 Hz. Subsequent preprocessing and analysis were performed offline using EEGLAB (v14.0, Delorme & Makeig, 2004), and FASTER (Nolan, Whelan, & Reilly, 2010) toolboxes. Bad channels were removed manually based on visual inspection. An IIR Butterworth .1–40 Hz band-pass filter (24 dB/oct, Order 8) (Luck, 2014), was used to remove excessive low drift and high frequency spike artifacts from the data. The resulting continuous data were segmented into epochs ranging from 200 msec prior to and 1000 msec after onset of the target picture and were baseline corrected to the prestimulus interval. Channel pop artifacts in individual epochs were automatically corrected with the FASTER toolbox. In combination with visual inspection, remaining bad epochs were marked using automatic algorithms for detecting improbable data (single channel, 6 SD; all-channels 4 SD). The average percentage of trials rejected for EEG artifacts and/or eye movements/blinks was 16.9% with a range of 3.7–25.9%. The range of the average rejections in each of the nine conditions (3 task × 3 picture type) was 15.4–18.3%. Pruned segments were re-referenced to the average reference. ICA was performed on the segmented data and artefactual independent components were manually rejected. Bad channels were interpolated and pre-processed epochs were then grand averaged.

Baseline ERPs were subtracted from negative and neutral waveforms to separate the ERP components of interest from activity elicited by the periodic appearance of pictures in the RSVP sequence (see Vogel, Luck, & Shapiro, 1998). ERP components were measured by averaging waveforms from 3 to 6 sensors (identical to those used in Kennedy et al., 2014) and time windows measured using a collapsed localizer technique (Luck & Gaspelin, 2017) which measures half amplitude latencies of the maximum voltage from an average waveform of all conditions. Measured components were analyzed using repeated measures, Analysis of Variance (ANOVAs) with Greenhouse-Geisser corrections. Following the ANOVAs, we performed post hoc comparisons on the group means using Fisher's Least Significant Difference (LSD) test. The LSD test does not involve a correction to the alpha level for multiple comparisons, because there is no inflation of family-wise

error rates for the special case of three conditions, as long as post-hoc tests are preceded by a significant main effect (Cardinal & Aitken, 2013).

2.2. Behavioral results

2.2.1. Picture detection

Picture detection performance was close to ceiling in all conditions (d′ values ranging from 3.83 to 4.48) and were not statistically analyzed. Not surprisingly, detecting pictures containing animates that are embedded in a stream of landscape pictures is an easy task, at last in part, because they contain close-ups of objects as opposed to the background pictures that are wide-angle pictures of landscapes and cities. This physical difference between distractors and background pictures might allow distractors to "pop-out" and capture attention automatically, a possibility that is addressed in Experiment 2.

2.2.2. Recognition memory accuracy

Fig. 2 shows recognition memory accuracy for negative and neutral pictures as a function of task. Recognition accuracy was lower in the Dual-Task condition compared to the Picture-Only task and surprisingly, was higher for neutral compared to negative pictures. Generalized linear mixed modelling (GLMM) using a binomial distribution with a logit link was applied to the percentage accuracy data for within-factors of image valence (Negative *vs* Neutral) and task (Picture-Only *vs* Dual-Task). There were significant main effects of task, $X^2(1, N = 4277) = 135.03$, $p < .001$, $\Phi = .177$, and image valence $X^2(1, N = 4277) = 46.26$, $p < .001$, $\Phi = .104$. The two-way interaction was not significant, $X^2(1, N = 4277) < 1$, $\Phi = .002$. The higher accuracy for neutral than negative pictures is surprising because one might have assumed that negative pictures would be more likely to intrude into awareness and be remembered compared to neutrals. However, the difference between these two picture categories may potentially reflect greater homogeneity of the negative pictures compared to neutrals which would make the recognition memory task more difficult for negative compared to neutral pictures. This finding, although puzzling, isn't critical to our analysis of automaticity.

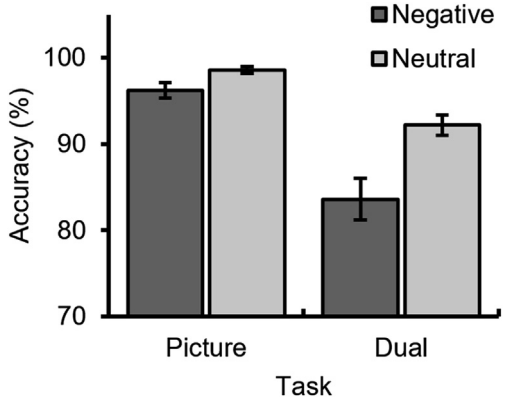

Fig. 2 – Picture recognition accuracy in Picture-Only and Dual-Task conditions. Error bars represent standard error the mean.

2.2.3. MOT accuracy

A correct response on the MOT task was defined as correctly selecting all three targets. Fig. 3 shows tracking accuracy as a function of task and image valence. Generalized linear mixed modelling using a binomial distribution with a logit link was applied to these data with factors of picture type (Negative, Neutral, & Baseline) and task (MOT-Only *vs* Dual-Task). There was a significant main effect of task, $X^2(1, N = 6405) = 26.14$, $p < .001$, $\Phi = .064$. Neither the main effect of image valence, $X^2(2, N = 6405) = 3.74$, $p = .15$, $\Phi = .024$, nor its interaction with task were significant, $X^2(2, N = 6405) < 1$, $\Phi = .005$.

These data show that MOT performance was worse in the Dual-task condition than the MOT-Only condition. Surprisingly though, the amount of interference was the same for all three picture types (negative, neutral, and baseline). One might have expected that tracking accuracy would be worse when the critical picture was an animate compared to a baseline picture because, in the Dual-task condition, both negative and neutral pictures require the participant to attend to the picture and store its identity in working memory in order to perform the recognition task at the end of the trial, whereas this operation is not required in the baseline condition. In addition, in the MOT-Only condition, one might have expected that attention capture by the negative or neutral pictures, due to their physical salience, would have interfered with tracking. And interference might have been larger for negative compared to neutral pictures if emotional pictures capture more attention than neutral pictures. The failure of these predictions to match the data suggests that MOT performance is relatively insensitive to interruptions, either due to automatic attention capture (as might occur in the MOT-Only condition) or to the use of higher–level processes involved in working memory consolidation, as would be required in the dual-task condition. This possibility is examined in Experiment 2.

2.3. ERP results

2.3.1. N2/EPN component

Our main question in this section is whether the amplitude of the N2s elicited by the animate pictures would decrease when participants were instructed to ignore the pictures and invest all of their attention into tracking objects. We measured the

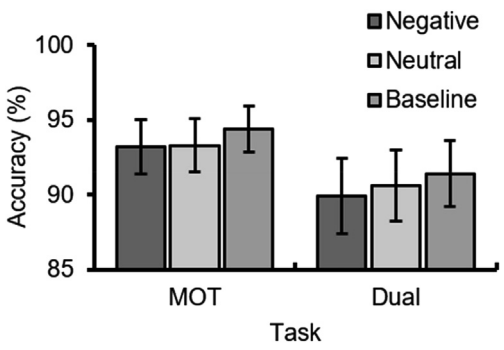

Fig. 3 – MOT accuracy for correctly reporting all three targets in the MOT-Only and Dual-Task conditions. Error bars represent standard error of the mean.

N2 components elicited by negative and neutral pictures and combined them to obtain the EPN (defined as the Negative–Neutral subtraction) from three contiguous sensors over the posterior left (EGI sensors: 64, 68, 69; approximately between T5 and O1 in the 10–20 system) and right (89, 94, 95; approximately between T6 and O2 in the 10–20 system) hemispheres. These sensor locations are the same as those used by Kennedy et al. (2014). The N2 is clearly visible in the Picture-Only condition in Fig. 4 and is larger for negative compared to neutral pictures. Its peak latency (230 msec) and posterior topography agree with previous results (e.g., Kennedy et al., 2014). Interestingly, the amplitude of the N2 appears to progressively diminish as attention is withdrawn from the picture stream in the Dual-Task and MOT-Only conditions. This apparent decrease in amplitude though is coincident with a positivity that has a scalp topography similar to the N2. This positivity may be the P_D component that is thought to reflect suppression of attention capture (Hickey et al., 2009; Sawaki & Luck, 2010), although this needs to be confirmed using lateralized stimulus presentations which would reveal whether it is larger over the contralateral hemisphere as is true of the P_D component.

The amplitude of the N2 was quantified as the average voltage between the half maximum amplitude points on either side of the peak (Luck, 2014) in a time window extending from 185 to 275 msec after onset of the animate picture (see Kennedy et al., 2014). These values were analyzed in a three-way repeated measures ANOVA using factors of valence (Negative *vs* Neutral), hemisphere (Left *vs* Right) and task (Dual-Task, MOT-Only, & Picture-Only). There was no significant main effect of hemisphere $F(1, 17) < 1$, $\eta^2p = .009$. However, there were significant main effects of valence, $F(1, 17) = 24.851$, $p < .001$, $\eta^2p = .594$ and task, $F(2, 34) = 10.130$, $p < .001$, $\eta^2p = .373$. An LSD pairwise test revealed significant differences in Picture-Only versus MOT-Only ($p = .002$), as well as MOT-Only versus Dual-Task ($p = .001$). The difference between Picture-Only and Dual-Task was not significant, ($p = .158$). In addition, the two-way interaction of hemisphere by valence was significant $F(1, 17) = 7.239$, $p = .015$, $\eta^2p = .299$, reflecting the larger difference in N2 amplitude between negative and neutral pictures in the right hemisphere compared to the left. Two-way interactions of hemisphere by task, $F(2, 34) < 1$, $\eta^2p = .042$, and valence by task, $F(2, 34) < 1$, $\eta^2p = .101$, were not significant. The three-way interaction between image valance, hemisphere, and task did not reach significance, $F(2, 34) < 1$, $\eta^2p = .006$.

We attempted to isolate the EPN elicited by the animate distractor pictures from the overlapping P_D component by subtracting the neutral condition from the negative. If the P_D is based on the physical salience of the animate pictures relative to the background pictures, it should be the same for negative and neutral pictures which both contain close-ups of animates compared to the background pictures which consist of wide-angle views of landscapes and cityscapes. If so, subtracting neutral from negative ERPs should eliminate the positivity and reveal the EPN. The EPNs for the three different tasks are shown in Fig. 5 and appear to be remarkably similar.

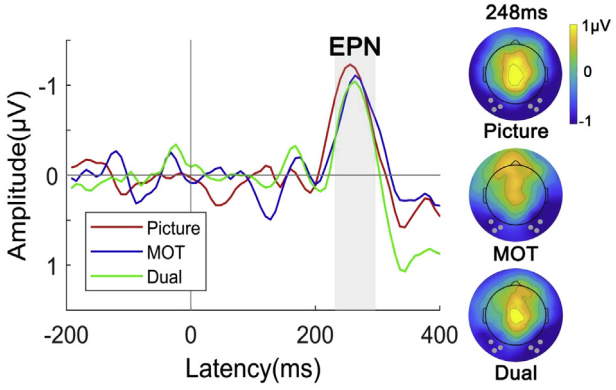

Fig. 5 – The EPN, time-locked to the presentation of the animate pictures. Shaded region (232–296 msec) represents the time window for measuring mean EPN amplitude. The topographic maps (right) show the scalp distribution of the EPN across tasks.

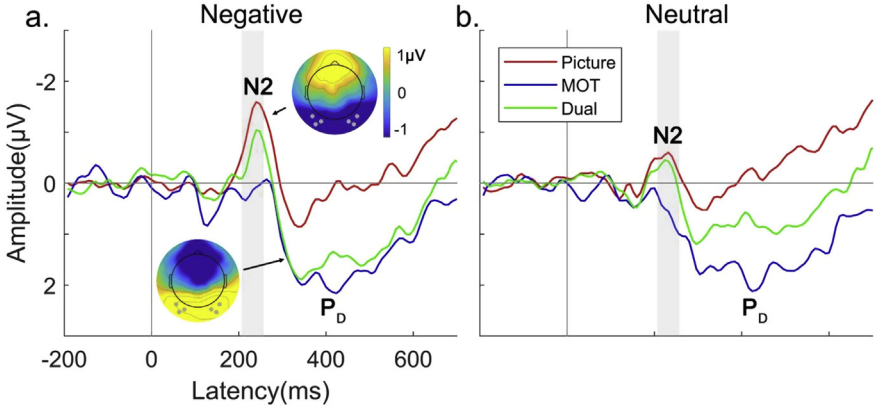

Fig. 4 – N2 and P_D components elicited by negative (a) and neutral pictures (b). The shaded region (185–275 msec) represents the time window used for calculating the mean amplitude of the N2 component. The upper topographic map corresponds to the N2 with a latency of 240 msec after the onset of the negative picture in the Picture-Only condition. The lower topographic map displays the P_D component at 416 msec after the onset of the negative picture in the MOT condition. Grey dots on the topographic map represent the sensors used for measurement. Note that the positivity we labelled as P_D appears to start earlier than this time window; early enough to affect the N2 component.

The peak latency of the subtraction curve was not the same as the peak of the "raw" N2 observed in the Picture-Only condition, so the measurement window was adjusted accordingly to 232–296 msec. We conducted a two factor repeated measures ANOVA on the mean amplitude of the subtraction curves using hemisphere (Right *vs* Left) and task (Picture-Only, MOT-Only, & Dual-Task) as independent variables. There was a main effect of hemisphere, $F(1, 17) = 8.696$, $p = .009$, $\eta^2p = .338$ reflecting a larger EPN in the right hemisphere. The main effect of task, $F(2, 34) < 1$, $\eta^2p = .033$ and its interaction with hemisphere, $F(2, 34) < 1$, $\eta^2p = .012$, were not significant. In addition, we performed one sample *t*-tests comparing the EPN amplitude to zero for each task condition for both hemispheres to determine if the EPN was present for all conditions. The *t*-test results show significant EPN activation for all three tasks and both hemispheres (all *p*s < .05).

To examine the strength of evidence favoring the null hypothesis for the effect of task (Picture-only, MOT-only, and Dual) on EPN amplitude, we conducted a Bayesian repeated-measures ANOVA on EPN amplitude using JASP (JASP Team, 2018), with default priors (Rouder, Morey, Speckman, & Province, 2012). This analysis produces a Bayes Factor (BF) consisting of the ratio of evidence in favor of the null compared to the alternative hypothesis. The BF_{01} (inverse BF) statistic, used in this analysis, favors the null model when it is greater than 1 with higher BF_{01} values indicating stronger evidence in favor of the null model. Results ($BF_{01} = 5.178$) indicate that the current findings are approximately 5 times more likely to be observed under the null model compared to the alternative model, providing "substantial" evidence (Jeffreys, 1961) that there are no differences in EPN amplitude across the three tasks.

In summary, these results show that negative pictures elicited a larger N2 compared to neutral pictures (the EPN), replicating previous work, and this difference was not affected by the amount of attention allocated to the picture stream (see Holmes, Kiss, & Eimer, 2006; Holmes, Nielsen, Tipper & Green, 2009, for similar findings). Assuming that the N2 is a measure of attention capture, this result is consistent with the idea that the negative and neutral distractors captured attention automatically. We also observed a prominent positivity in response to the salient animate distractor pictures in conditions where participants were attempting to allocate all (MOT-Only) or most (Dual-Task) of their attention to the MOT task. This component had a scalp topography similar to that of the N2 and made it appear as if the N2 had been reduced or eliminated in conditions in which participants were at least partially ignoring the pictures (MOT-Only & Dual-Task). This may be the P_D component which is thought to represent suppression of attention capture by physically salient stimuli (Sawaki et al., 2012) but additional research is required to confirm this. In any case, these results are puzzling. The presence of an EPN in the MOT-Only condition might be taken as evidence that emotional distractors automatically captured attention. But if this was the case, it surprisingly didn't have any effect on tracking accuracy. Alternatively, perhaps attention capture was suppressed which is in line with the presence of a P_D component in both conditions in which participants had to track objects (Gaspelin & Luck, 2019). But if capture was suppressed why did we observe an EPN component whose amplitude was the same across all three task conditions? We consider these puzzling findings in the discussion.

2.3.2. *P3b component*

The P3b component was isolated from the background RSVP activity by subtracting the baseline condition from the negative and neutral conditions. We used the same six EGI sensors (54, 55, 61, 62, 78, 79; approximately between Cz and Pz in the 10–20 system) used by Kennedy et al. (2014) in their analysis of the P3b. A time-window of 400–992 msec was selected using a collapsed localizer technique (Luck & Gaspelin, 2017) by averaging the P3b component across all conditions and measuring the half amplitude of the maximum positive value. The waveforms in Fig. 6 depict P3b activity elicited by the negative and neutral animate pictures for the three tasks.

Two clear findings are visible in Fig. 6: The P3b is larger for negative than neutral pictures and it gets smaller as more attention is withdrawn from the pictures and allocated to the tracking task. A two-factor repeated measures ANOVA with factors of image valence (Negative *vs* Neutral) and task (MOT-Only, Dual-Task, & Picture-Only) revealed significant main effects of task $F(2, 34) = 16.232$, $p < .001$, $\eta^2p = .488$ and valence $F(1, 17) = 11.530$, $p = .003$, $\eta^2p = .404$, as well as their interaction, $F(2, 34) = 11.348$, $p < .001$, $\eta^2p = .400$. Pairwise LSD tests on the main effect of task revealed that P3b amplitude was significantly different for all three pairs of tasks: Picture-Only versus Dual-Task ($p < .001$), Picture-Only versus MOT-Only ($p < .001$) and Dual-Task versus MOT-Only ($p = .009$). The main effect of task shows that the P3b amplitude elicited by the animate pictures decreased as progressively less attention was allocated to the picture stream. The significant effect of valence reflects the greater P3b amplitude for negative pictures compared to neutrals which is a consistent finding in the literature (e.g., Foti, Hajcak, & Dien, 2009; Kennedy et al., 2014).

To examine the task by valence interaction from the above analysis in more detail we created Negative–Neutral difference scores (shown in Fig. 7) and examined how they were affected by the three tasks. We used a measurement window of 304–528 msec based on the average half amplitude of the P3b difference between Negative and Neutral across the three task conditions. These amplitudes were analyzed in a one-way repeated-measures ANOVA. Once again, there was a significant main effect of task, $F(2, 34) = 17.422$, $p < .001$, $\eta^2p = .506$. LSD comparisons revealed that differences between all three task pairs were significant ($p < .05$). This shows a clear effect of attention on the P3b elicited by the animate pictures. As attention was withdrawn from the picture stream, the effect of valence on the amplitude of the P3b component was reduced.

We also performed single-sample *t*-tests to determine if the difference curves in Fig. 7 were significantly different from zero in each task condition. We found significant P3b components for all three tasks (all *p*s < .05) showing that even in the MOT-Only condition, negative pictures elicited slightly more P3b activity than neutrals.

Overall, the results indicate that negative pictures elicited larger P3b components than neutral pictures, replicating prior research (Foti et al., 2009; Kennedy et al., 2014). In addition, the effect of valence on P3b amplitude was modulated by the

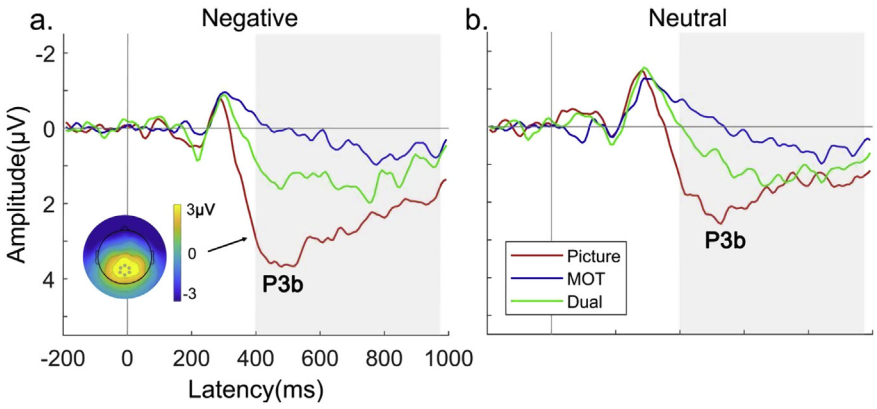

Fig. 6 – The P3b component elicited by negative (a) and neutral (b) animate pictures. Shaded region (400–992 msec) represents the time window for measuring mean P3b amplitude. The topographic map displays the P3b component at 512 msec after the onset of the negative picture in the Picture-Only condition.

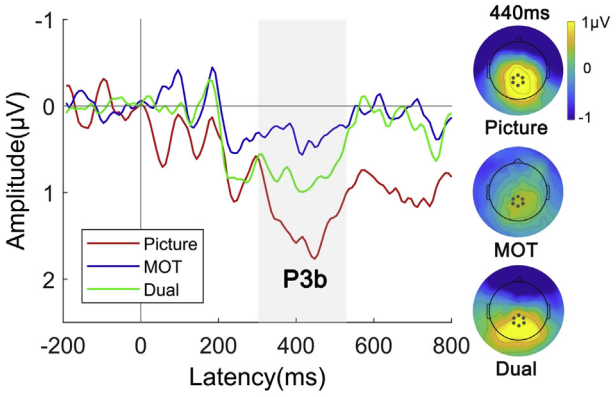

Fig. 7 – The negative–neutral difference curve (left) as a function of task reveals the effect of valence on the P3b component. Shaded region (304–528 msec) represents the time window for measuring mean P3b amplitude. The topographic maps (right) display the P3b component for each task. Grey dots in the topo plot indicate the relevant sensors.

amount of attention directed to the picture stream, becoming smaller as attention was withdrawn from the picture stream and allocated to the MOT task. This result stands in sharp contrast to the EPN component whose amplitude was independent of attention.

2.4. *Discussion*

The results of Experiment 1 show that several aspects of processing of emotional pictures are affected by how much attention is allocated to them, showing that at least some aspects of emotional processing are not automatic. First, recognition memory for both neutral and emotional pictures was reduced when attention was shared between the picture recognition and the MOT task. Consistent with this behavioral cost, the amplitude of the P3b component elicited by emotional pictures was strongly reduced in dual-task conditions and virtually eliminated when participants were instructed to ignore the picture stream and allocate all their attention to the tracking task. However, even in this condition there was a small but significant P3b that was larger for negative than neutral pictures, suggesting that participants were not completely successful in ignoring the emotional picture.

Better evidence for automatic attentional capture would be a decrease in tracking accuracy in the MOT-Only condition when an emotional picture appeared in the stream. If the emotional picture captures attention, one might expect tracking accuracy to suffer as previous evidence shows that MOT depends on allocation of visual attention to the tracked objects (Tran & Hoffman, 2016).

However, behavioral evidence for automatic attention capture by emotional pictures was surprisingly absent as tracking accuracy was unaffected by the appearance of emotional pictures in the stream. This was the case even though we observed equivalent EPN components (presumably reflecting attention capture) in all three task conditions. However, the failure of the tracking task to reveal any attentional capture effects may reflect a lack of sensitivity of MOT performance to *relatively brief* withdrawals of attention from the tracked objects. For example, Alvarez, Horowitz, Arsenio, Dimase, and Wolfe (2005) found that moving objects in MOT could be "turned off" (invisible while still moving) for periods of nearly one third of a second without any loss in tracking accuracy. This leaves open the possibility that negative and/or neutral pictures did capture attention during the MOT task, but the capture period was within the time window in which MOT is insensitive to the absence of attention.

The second experiment was designed to test the possibility that emotional pictures briefly capture attention even when participants are trying to ignore them in order to perform the MOT task. Participants were required to monitor the tracked objects for the occurrence of a gap and make speeded responses indicating whether the gap was on the left or right side of the object. If emotional pictures capture more attention than neutrals, even for a brief time, the reaction time (RT) of the response to the gap should be slower in the presence of negative compared to neutral distractors. On the other hand, if initial capture by distractors is driven solely by physical

salience, which is comparable for both distractor types, negative and neutral pictures should produce equivalent delays in gap RT. We also examined whether a delay in RT to the gap was accompanied by a delay in the N2 elicited by the gap which would support the claim that RT delays are being driven by *competition for attention* rather than effects at later processing stages, such as response selection. We didn't examine the P3b elicited by distractors in this study because the gap task elicited prominent P3b activity which obscured any P3bs elicited by the distractors.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Twenty-two right-handed, neurologically normal participants were recruited through a university-sponsored classified ad. Two participants were excluded for excessive artifacts in the ERP recording (more than 30% of total epochs), resulting in 20 participants (9 women, 13 men, mean age: 22.7, age range: 19–32). All participants were right handed and had normal or corrected-to-normal vision. The University of Delaware Institutional Review Board approved the study and all participants provided informed consent and were compensated at the rate of $10 per hour for their participation.

3.1.2. Stimuli

A different set of animate and landscape pictures was used for Experiment 2. The landscape pictures were drawn randomly from a set of 122 landscape and architectural photographs (see Fig. 1). There were 38 negative (violence, medical trauma, threatening animals) and 38 neutral (people, animals) valence animate pictures that were presented as distractors within the RSVP stream. Similar to Experiment 1, some pictures were taken from the IAPS and others were pictures gathered from the internet. Dependent measures t-tests revealed that the valence of the negative pictures ($M = 8.08$, $SE = .06$) was significantly more negative, $t(74) = 32.222$, $p < .001$, than the neutrals ($M = 4.65$, $SE = .09$). Negative pictures ($M = 6.28$, $SE = .05$) also had significantly higher arousal, $t(74) = 27.723$, $p < .001$, compared to neutrals ($M = 3.98$, $SE = .09$). Other stimuli dimensions were identical to Experiment 1 except that a target in the form of a small gap (.48°) was presented on either the right or left side of one of the tracked disks.

3.1.3. Procedure

The procedure of Experiment 2 was similar to the MOT-Only condition of Experiment 1 except that a gap could appear in the left or right side of one of the tracked circles. Participants were instructed to track two moving objects and report the orientation of the gap by making a speeded button press with the left or right mouse button to indicate which side of the object contained the gap. The gap appeared either simultaneously with the distractor (0 msec SOA), 300 msec after distractor onset (300 msec SOA), or not at all (No-gap). Participants were instructed to ignore the RSVP sequence and they were never queried about its contents. The duration of the gap was 500 msec which was long enough to ensure high accuracy, allowing us to use reaction time (RT) as the principal dependent variable reflecting attentional capture.

The experiment consisted of a factorial combination of 3 distractor types (Negative, Neutral, & Baseline), 3 SOAs (0 msec, 300 msec, & No-gap), 2 gap-types (Left & Right), and 2 distractor locations in the stream (4th & 8th). The combination was repeated 16 times for a total of 576 trials. For analysis, trials were combined across distractor locations and gap type providing 64 trials in each of the nine relevant categories (3 SOAs × 3 distractors). The entire set of trials was presented in a pseudo-random order. At the end of the trial, following the tracking response and its associated accuracy feedback, participants received feedback on the speed and accuracy of the gap detection response. Feedback consisted of the following types: correct response ('Correct'), incorrect response ('Incorrect'), False Alarm ('No gap was presented)', and Miss ('Failed to make response'). Participants were also given feedback on whether they had moved their eyes or blinked.

At the beginning of the study, participants initially received 16 practice trials with just the tracking task without gap detection, and later received 32 more with both tracking and gap detection tasks. During each break, participants were provided with cumulative tracking accuracy, average gap reaction time, average gap accuracy, and number of eye blink/movements. If MOT accuracy failed to reach 50% or exceeded 90% during the previous block, the speed of the moving objects was adjusted using a staircase procedure (i.e., speed was lowered after 1 incorrect response and increased after 3 correct responses) during the break to meet 80% accuracy. Participants were debriefed and paid at the end of the experiment.

3.1.4. EEG recording and analysis

The EEG recording, and pre-processing procedure was identical to that used in Experiment 1. The average percentage of trials rejected for EEG artifacts and/or eye movements/blinks was 19% with a range of 8.5–28.8%. The range of the average rejections in each of the nine conditions (3 SOA × 3 picture type) was 16.4–20.5%. In Experiment 2, we were interested in the EPN components elicited by emotional pictures as well as the N2 component elicited by the gaps. The EPN served as a confirmation that the task-irrelevant emotional distractor pictures capture attention. The gap N2 was of interest because it should reflect the competition for attention between the salient distractor picture and the task-relevant gap. In the 0 msec SOA condition, the EPN and Gap N2s occurred close in time and produced overlapping activity. These components were isolated from each other and from other overlapping ERP components, such as the P_D, using various subtractions. The nature of these subtractions depended on a simple model of how the overlapping components were related to each other. We assumed first that the EPN component was not affected by the requirement to discriminate the gap. In other words, even when the gap occurred simultaneously with the distractor picture, attention was allocated to the distractor picture first

and any effects of competition for attention were assumed to only affect the gap N2. This is reasonable given that the distractor picture is a new, object appearing in the stream of background pictures and should be much more salient than the appearance of the gap which represents a subtle change in a pre-existing object (Yantis & Hillstrom, 1994). In addition, this assumption makes a strong prediction which can be used to assess its validity; namely, the EPNs resulting from different subtractions should all be identical.

Second, we assume that negative and neutral distractors will have the same effect on the N2 elicited by a simultaneous gap. The effect of attention being initially captured by a distractor picture should result in a delay in both the overt response and the allocation of attention to the gap, particularly because the gap had a long duration allowing for correct discrimination of its orientation even when attention is delayed. The claim of a similar delay for both distractor types (negative & neutral) rests on the assumption that it is the *physical salience* of the distractor picture that is driving attention capture and physical salience should be the same because negative and neutral pictures are similar to each other (both contain close-up views of people and animals) and therefore are equally dissimilar from the background pictures (wide-angle views of landscapes and cityscapes). This claim is important because it holds that *emotional valence is not playing any role in the initial capture of attention by a negative distractor picture*. This assumption predicts that the Gap N2s associated with negative and neutral distractors should be identical and delayed by the same amount.

As an example, consider the case of negative and neutral pictures appearing simultaneously with a gap. The EPN component can be isolated using the negative-neutral subtraction. The P_D component associated with both distractors should be eliminated by this subtraction as it was in Experiment 1. In addition, the N2s elicited by the gap should be the same for both distractors because the distractors have the same physical salience and should delay the gap N2 by the same amount. Therefore, this component should be eliminated in the subtraction as well. If these assumptions hold, the result should be an EPN component that is the same as that observed in the No-gap condition where there is no overlapping Gap N2. It should also be the same as the EPN observed in the 300 msec SOA condition where there is no overlapping gap N2 because the gap appears 300 msec after the distractor picture.

The gap N2 can be isolated in the 0 msec SOA condition as follows: subtracting the Negative No-gap ERP from the Negative plus gap ERP should isolate the gap N2 because the Negative ERP (and associated P_D) will be the same in both conditions and will therefore be eliminated in the subtraction. The same should be true for the neutral distractor condition. The assumptions stated above predict that the resulting gap N2 should be the same for negative and neutral distractor conditions. In addition, the gap N2s can be compared to the RT data to see if a similar pattern emerges. For example, if the RTs to the gap are delayed by similar amounts in the presence of negative and neutral distractors, we might expect the same to hold true for the latency of the gap N2s resulting from the two subtractions (assuming that the gap RT reflects a delay in allocating attention to the gap because of prior attention capture by the distractor).

3.2. Behavioral results

3.2.1. Gap discrimination reaction time and accuracy

As expected, the long gap presentation duration (500 msec) resulted in accuracy that was close to ceiling ($M = 95\%$, $SE = .003$) allowing us to focus on RT as a measure of capture. Outliers that exceeded ±2 standard deviations per condition within individuals were rejected (Lachaud & Renaud, 2011). Reaction time (RT) to the gap is shown in Fig. 8. It reveals that RT was delayed relative to the baseline condition (no distractor picture) when the gap appeared at the same time as the distractor picture (0 msec SOA). The delay was approximately the same for negative and neutral distractor pictures. At 300 msec SOA, the RT delay was eliminated. Statistical analysis confirms this description. A Generalized Linear Mixed Models (GLMM) procedure using an inverse Gaussian distribution with an identity link (Lo & Andrews, 2015) was applied to analyze two within-factors of distractor type (Negative, Neutral, & Baseline) and SOA (0 & 300 msec). There was a significant main effect of distractor type, $X^2(2, N = 6065) = 46.04$, $\Phi = .046$, $p < .001$., and SOA, $X^2(1, N = 6065) = 7.25$, $\Phi = .035$, $p = .007$. The two-way interaction was also significant $X^2(2, N = 6065) = 29.30$, $\Phi = .07$, $p < .001$. LSD pairwise contrasts revealed that when the SOA was 0 msec, RTs in the negative and neutral conditions significantly differed from baseline by approximately 30 msec ($p < .001$). The difference between negative and neutral was not significant ($p = .999$). There were no RT differences within the 300 msec SOA conditions ($p > .05$). Overall, responses to gaps were significantly slowed by the simultaneous presentation of a negative or neutral distractor picture but this interference was short-lived and was eliminated if the gaps occurred 300 msec after the distractor picture.

To examine the strength of the null main effect of valence (Negative *vs* Neutral) on gap reaction time, we conducted separate Bayesian repeated-measures *t*-tests on negative and neutral distractors for 0 msec and 300 msec SOA. Results for the 0 msec SOA ($BF_{01} = 4.268$) and the 300 msec SOA

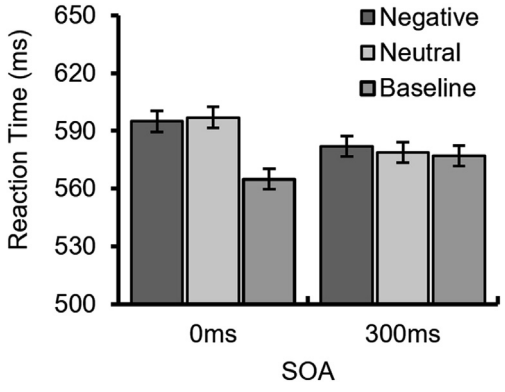

Fig. 8 – Reaction time for discriminating the orientation of the gaps presented at SOAs of 0 and 300 msec. Error bars represent standard error of the mean.

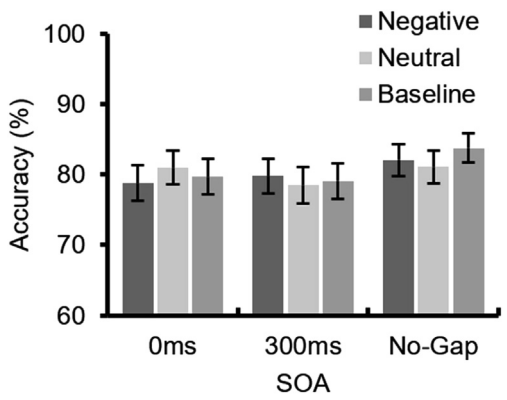

Fig. 9 – MOT tracking accuracy for reporting both tracked objects. Error bars represent standard error from the mean.

(BF_{01} = 3.638) indicate that the current findings are approximately 3–4 times more likely to be observed under the null model compared to the alternative model, providing substantial evidence that there are no differences in gap RT between negative and neutral conditions.

3.2.2. *MOT accuracy*

MOT accuracy (shown in Fig. 9) was assessed through GLMM using a binomial distribution with a logit link. Differences in within-subject factors of distractor type (Negative, Neutral, & Baseline) and SOA (0 msec, 300 msec, & No-Gap) were tested. There was a significant main effect of SOA, $X^2(2, N = 10837) = 13.57$, $\Phi = .035$, $p = .001$. Tukey pairwise comparisons revealed a significant difference between the No-gap and both the 0 msec ($p = .018$) and the 300 msec condition ($p = .001$). Neither the main effect of distractor type $X^2(2, N = 10837) < 1$, nor the two-way interaction $X^2(4, N = 10837) = 4.97$, $\Phi = .021$, $p = .29$ reached significance. We don't consider the SOA effect here to be important. The main finding is that these results replicate those of Experiment 1 in showing that MOT accuracy is not affected by the appearance of negative or neutral distractors in the to-be-ignored picture stream even though these distractors briefly captured attention as shown by the delay in gap RT (Fig. 8).

3.3. *ERP results*

3.3.1. *Gap N2 latency results*

The N2 components elicited by the gap are shown in Fig. 10 which shows that the N2 elicited by the gap is delayed by about the same amount when it appears simultaneous with a negative or neutral distractor. When the gap appears 300 msec after the distractor, the delay is eliminated. This pattern is similar to that observed for gap RT and suggests that the RT delays reflects a delay in *attending* to the gap. The latency of the N2 component elicited by the gap was defined as the 50% fractional area latency (Luck, 2014). The waveforms were measured at the same six bilateral, posterior sensors (EGI sensors: 58, 64, 65, 90, 95, 96) used in the first experiment. A measurement window of 150–500 msec after the onset of the gap was used to measure gap N2 latency for both 0 msec and 300 msec conditions. The same GLMM procedure used in analyzing the behavioral RT results was applied. The test revealed a significant main effect of valence, $X^2(2, N = 120) = 21.96$, $\Phi = .43$, $p < .001$. The main effect of SOA was not significant, $X^2(1, N = 120) < 1$, $\Phi = .08$. The two-way interaction was significant, $X^2(2, N = 120) = 6.18$, $\Phi = .23$, $p = .05$. Tukey pairwise comparisons revealed that when the gaps were simultaneously presented with the distractor pictures (0 msec SOA), the gap N2s were significantly delayed in the negative ($M = 340$ msec, $SE = 13.3$, $p = .003$) and neutral ($M = 356$ msec, $SE = 13.7$, $p < .001$) conditions relative to baseline ($M = 297$ msec, $SE = 12.3$). There was no significant difference between negative and neutral ($p = .8$). In the 300 msec SOA condition, there was no differences between negative ($M = 333$ msec, $SE = 13.1$), neutral ($M = 330$ msec, $SE = 13.0$), and baseline ($M = 312$ msec, $SE = 12.6$), (all p's > .05).

Bayesian repeated-measures *t*-tests on negative and neutral distractors for 0 msec and 300 msec SOA, were performed to test the strength of evidence for the absence of any difference between negative and neutral distractors in the N2 delay effect. Inverted Bayes factors for 0 msec SOA

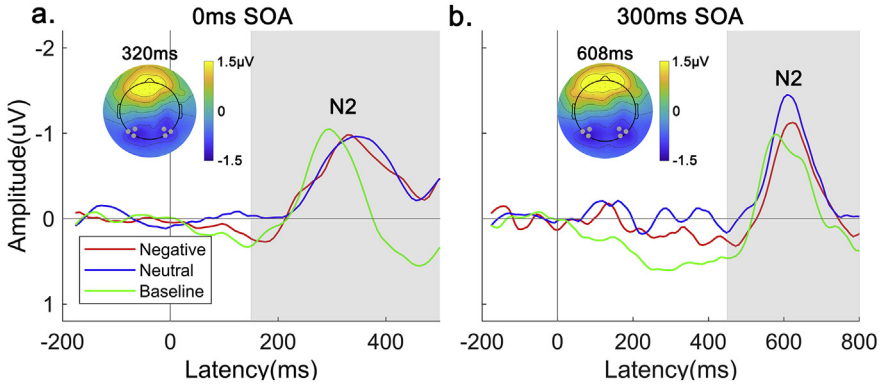

Fig. 10 – N2 components elicited by gap presentations at 0 msec (a) and 300 msec SOA (b) time locked to distractor picture presentation. The topographic maps correspond to the N2 at 320 msec (a) and 608 msec (b) in the Negative picture condition. Grey dots represent the electrodes used for measuring the N2 component. Shaded regions represent the time windows used for measuring N2 latency. Waveforms have been smoothed with a 50 msec averaging window for display.

($BF_{01} = 2.94$) and the 300 msec SOA ($BF_{01} = 4.295$) indicate that the current findings are approximately 3–4 times more likely to be observed under the null model compared to the alternative model, providing substantial evidence that there are no differences in gap RT between negative and neutral conditions.

3.3.2. *Distractor EPN amplitude results*

Fig. 11a shows ERP components produced by the negative and neutral distractors minus the baseline waveforms for No-gap, 0 msec and 300 msec SOAs. Consistent with Experiment 1, to-be-ignored distractors elicited a prominent positive waveform peaking around 312 msec at posterior electrode sites. To isolate the EPN component, the neutral distractor waveform was subtracted from the negative for each SOA conditions (shown in Fig. 11b). As in Experiment 1, the mean amplitude of the EPN was measured from left (64, 68, & 69) and right (89, 94, & 95) posterior sensors. An averaging window of 240–312 msec was defined by the half peak amplitude of the three SOA waveforms. Again, a prominent EPN component was elicited, replicating the MOT-Only condition of Experiment 1 and other studies that combined MOT with task-irrelevant emotional distractors in the background (Kim, & Hoffman, in prep; Kim, Taylor, & Hoffman, 2017). Single-sample *t*-tests showed that the amplitude of the EPN at each SOA was significantly different from zero: 0 msec, $t(19) = -5.597$, $p < .001$; 300 ms $t(19) = -4.149$, $p = .001$; No-gap, $t(19) = -5.503$, $p < .001$. The EPNs are remarkably similar across SOAs and support our assumption that attention capture by the distractor pictures was not affected by the presence or absence of a gap. Repeated measures ANOVA revealed no significant effect of SOA on EPN amplitude, $F(2, 38) < 1$, $\eta^2 p = .017$.

3.4. *Discussion*

Experiment 2 shows that task-irrelevant, negative and neutral distractor pictures that appear while participants are performing a multiple object-tracking task (MOT) briefly capture attention. This attention capture is revealed in delayed RTs to gaps appearing in the tracked objects when the gaps appear simultaneously with the distractor picture. The delay effect is short lived as it disappears when the gaps appear 300 msec after the distractor picture. Importantly the size of this delay was the same for negative and neutral distractors suggesting that capture in based on physical salience and not emotional salience. A similar pattern was observed in the N2 component elicited by the gap which was delayed when the gap occurred simultaneously with the distractor picture. The delay was the same magnitude for negative and neutral distractor pictures and was eliminated with a 300 msec delay. The similarity of the pattern for RT and N2 latency suggests that the RT delay caused by the presentation of a simultaneous distractor picture is attributable to a delay in *attending to the gap* rather than a delay in some other processing stage such as response selection. In contrast, tracking accuracy was once again unaffected by the appearance of a negative or neutral distractor in the stream, in agreement with the results of Experiment 1. This shows that MOT performance is insensitive to brief interruptions in attention to the tracked objects.

These results strongly support the hypothesis raised in the introduction that the initial capture event in EIB is the same for negative and neutral distractors and is based on the *physical salience* of the distractors relative to the background pictures in the stream. Emotional salience appears to play no role in the initial capture but presumably exerts a powerful interference effect in downstream processes that follow the initial visual processing responsible for attentional capture. These later processes are reflected in the P3b component which is related to the size of the blink observed in the EIB paradigm. Negative distractors that elicit a P3b are more likely to produce a blink of a closely following target.

The basis for physical salience capture in EIB is illustrated in Fig. 12, which shows the average of all the pictures in the negative, neutral, and baseline categories. The baseline pictures, which are wide angle views of landscapes and cityscapes, reveal a strong "horizon line" with blue sky above and green texture below. In contrast, the distractor pictures lack the horizon line and instead have red colored texture concentrated in the center of the picture corresponding to the

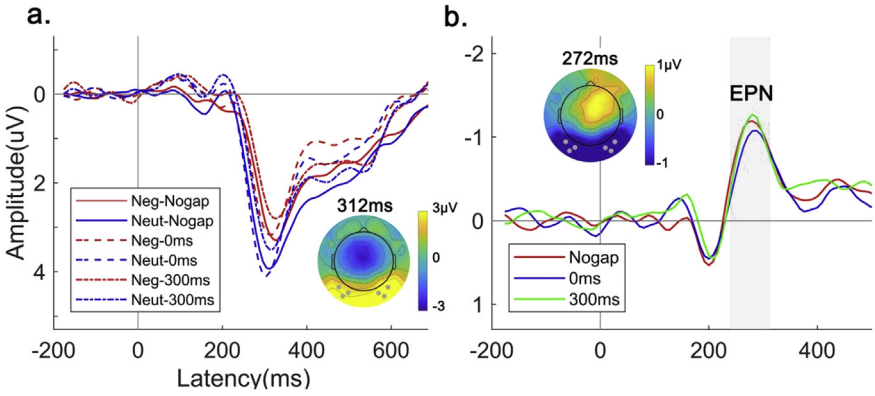

Fig. 11 – N2/P_D waveforms (a) and the EPN (b) components. The topographic map in (a) corresponds to the positivity at 312 msec after onset of the negative picture in the no-gap condition. The map in (b) shows the EPN at 272 msec in the no-gap condition. Grey dots represent the sensors used for measuring the components. The shaded regions represent the time windows for calculating the mean amplitude.

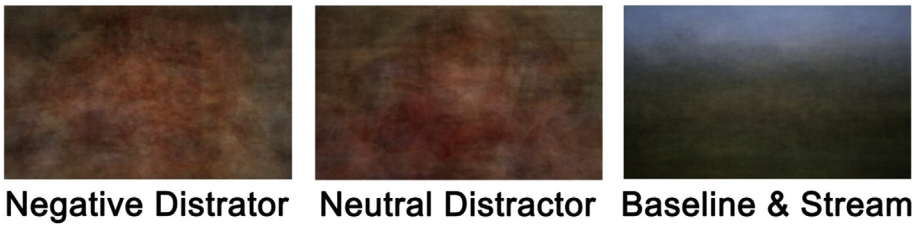

Fig. 12 – Averaged pictures of each distractor type. Baseline distractors are landscape images that are presented in the background stream. Negative and neutral distractors are physically similar to each other but considerably different from the stream pictures.

appearance of people and animals as central element in the picture. Negative and neutral distractors are similar to each other and different from the baseline pictures leading them to have equivalent physical salience-driven attentional capture relative to the baseline pictures appearing in the rest of the stream. Note that the use of a procedure in which the emotional pictures are always task-irrelevant, and observer's attention is occupied by a difficult, attention-absorbing task (MOT plus gap detection) at the moment of presentation of the emotional distractor eliminated any behavioral sign of emotional capture. However, we still observed an EPN component in which the N2 elicited by the emotional picture was larger than the neutral picture N2. If the amplitude of the N2 reflects amount of attention capture, as is often assumed, then we should conclude that the negative distractor captured more attention than the neutral distractor. Clearly, our behavioral and ERP measures of attention capture are in conflict. We address this puzzle in the general discussion.

4. General discussion

The present study used behavioral and ERP measures to investigate the automaticity of attentional capture by emotional pictures using an RSVP presentation paradigm and stimulus materials similar to those used in studies of emotion-induced blindness (EIB). Emotional capture is particularly robust in EIB, sometimes producing a 40% decrease in the ability to discriminate closely following target pictures. In addition, as McHugo, Olatunji, and Zald (2013) pointed out in their review of the research on EIB, there is no evidence that top-down attention can modulate the strong capture effects in this paradigm. Therefore, EIB appears to be something of a gold standard for the claim that attentional capture by emotional pictures is automatic. We set out test this claim and to understand the factors that make emotional capture in EIB so robust.

In the first experiment, participants carried out two tasks: tracking multiple moving objects (multiple object tracking or MOT) and identifying neutral and negative pictures embedded in a stream of background scene pictures, the same stimuli used in EIB. In different blocks of trials, participants performed each of the single tasks by themselves or together. We found that emotional distractors elicited the same EPN component in all three attention conditions, suggesting that this component was automatic. However, a later component elicited by the distractor pictures, the P3b, was strongly affected by attention, decreasing in amplitude as attention was withdrawn from the picture task and allocated to the tracking task. Surprisingly, though, tracking performance was not affected by emotional distractors appearing in the RSVP stream, even though these pictures elicited an EPN component which is thought to reflect attention capture.

One explanation for the failure to find behavioral evidence of attention capture is that the MOT task is insensitive to brief interruptions. We tested this idea in Experiment 2 in which participants were required to make speeded responses to gaps appearing in one of the tracked objects. Delays in the reaction time (RT) of the response to the gaps should be sensitive to even brief interruptions in attention to the tracked objects. We did observe small delays in RT when either a negative or neutral distractor appeared simultaneously with the gap. Similar delays occurred in the N2 component elicited by the gap showing that the RT delay was caused by competition for attention between the two tasks. These delays were the same magnitude for both distractors suggesting that they were caused by *physical salience* rather than *emotional salience*. An analysis of the pictures used in the EIB paradigm showed that negative and neutral distractors are similar to each other and different than the background pictures making it plausible that distractors initially capture attention in EIB based on physical salience.

If attentional capture in EIB is driven by physical salience, which is the same for both distractor types, why is the suppression of the following target (the blink) so much larger for emotional pictures than neutrals? The explanation may lie in later processes that are located downstream from the early visual processes responsible for capture. When negative distractors capture attention during EIB, they may proceed to later processes such as working memory consolidation and conscious awareness. In other words, attention serves as a key to unlock higher level processes that have strong capacity limits. Neutral pictures don't access these later processes perhaps because they are effectively suppressed shortly after capturing attention. In contrast, negative pictures often do proceed to later stages as reflected in the substantial P3b component that they elicit during EIB. Because these later stages are highly limited in capacity, the emotional picture blocks access by the target picture, leading to failures of awareness for the target and providing the basis for the blink

in EIB. The same process occurs in the AB and shows that EIB and AB are essentially mediated by the same mechanisms. It isn't clear though why emotional valence allows the picture to overcome suppression and gain access to higher level processes while a neutral picture is effectively suppressed. Perhaps emerging information about the meaning and content of the emotional picture reveals that it is important and should have access to working memory and awareness. This places emotional salience at later stages while physical salience is determined in early vision. Landman, Sharma, Sur, and Desimone (2014) reported similar results in the macaque monkey. Task-irrelevant emotional faces only had an interfering effect on a grating discrimination task when they had a duration of at least 200 msec.

According to the view, emotional salience emerges later than physical salience and is based on information associated with the attended object. Importantly, this information depends on attention and therefore cannot initially be the basis for capturing attention. This runs counter to claims that emotional valence is computed rapidly in specialized structures such as the amygdala which might allow for rapid attentional capture based on emotional salience, perhaps on a par with the speed of physical salience capture. The claim that EIB is based on physical rather than emotional salience predicts that removing the physical salience associated with distractors would also eliminate EIB because distractors would no longer capture attention and they would be unable to access the later processes that lead to emotional salience. This prediction was tested by Baker, Hoffman, and Turco (2018) who replaced the scene pictures in the EIB paradigm with pictures containing close-ups of people and animals. In other words, the baseline pictures were replaced with "neutral pictures" which are physically similar to the negative distractor pictures (see Fig. 12) but lack emotional valence. This resulted in a 75% reduction in the magnitude of EIB from 15% with the typical, dissimilar background pictures to 4% with similar pictures. This residual EIB effect might simply reflect less than perfect similarity between negative and baseline pictures. But in any event, it shows that physical salience plays a powerful role in the EIB effect.

One puzzling aspect of our results is that although there was no behavioral evidence of emotional capture in Experiment 2, we did observe an EPN component which is assumed to reflect emotional capture. This disconnect between behavioral and ERP measures may be more apparent than real, however, because there is emerging evidence that the N2pc, and possibly the EPN as well, do not directly reflect the capture process itself. First, it is important to acknowledge that we don't know yet whether the non-lateralized EPN and the lateralized N2pc are the same component as the relevant research establishing this connection has yet to be done. But we believe this is a reasonable tentative hypothesis worthy of exploration.

There are a number of reasons for assuming the N2pc is not a direct measure of attention shifting but rather a measure of processes that unfold after attention has been directed to an object. These processes may involve identification of the object and retrieval of information associated with it, such as reward history, etc. First, several studies support the claim that N2pc is not a measure of attention shifting. Kiss, Van Velzen, and Eimer (2008) pre-cued observers to attend to the likely location of a search target and found that the N2pc elicited by the target was very similar to the case where attention was not pre-cued. This shows that the N2pc is not a measure of the attention shifting process itself but reflects downstream effects of attending to an object (see also Zivony, Allon, Luria, & Lamy, 2018). In addition, they showed that when the pre-cued object was a target, the N2pc was larger than when it was a distractor. As the authors point out, "the N2pc triggered in response to pop-out visual search targets does not reflect processes involved in covert shifts of spatial attention, but is instead linked to spatially selective attentional mechanisms that occur after such shifts are completed" (p. 248) and the N2pc "may not only be indicative of attentional target selection, but also, at least to some degree, reflects the spatially specific processing of potentially task-relevant features" (p. 248). Similarly, Theeuwes (2010, p. 86) concluded "the occurrence of the N2pc does not say anything about attentional capture but about the post-selection processing occurring at a particular location." In other words, a larger N2pc component is related to attention in the sense that it is primarily attended objects that elicit this component. However, the amplitude of the N2pc also reflects properties of the attended object itself such as whether or not it is a target. Given that the N2pc is also larger for objects associated with reward (Kiss, Driver, & Eimer, 2009) it may not be surprising that it would be larger for emotionally valenced pictures as well.

In the current experiments, the emotional picture always occurs in an attended location because the observer is attending to a single, centrally located stream of pictures. We conjecture that the EPN reflects an automatic identification of all objects occurring within the window of spatial attention and it is this process that elicits the EPN. It doesn't produce or depend on attention capture as attention is already allocated to the location of the picture stream as part of the object tracking task. Consider the following study which makes this point clear. Kim and Hoffman (in prep) replicated the current Experiment 2 in which the distractor pictures are dissimilar from the background pictures and therefore pop-out of the stream. Both distractors produced a 40 msec delay in RTs to the gap, consistent with the present results. They also ran a second experiment in which the physical salience of the negative and neutral distractor pictures was reduced by using background pictures that were similar to the neutral pictures. In this case, the distractor induced delay in RT to the gap was largely eliminated (5 msec for the negative distractor and approximately 0 for the neutral distractor). However, the EPN associated with negative distractors was the same magnitude in both experiments. These results show that the ability of distractors to capture attention in EIB is largely eliminated when their physical salience is reduced and the same EPN component is elicited by the distractors regardless of whether or not there is attentional capture. This provides strong evidence that the N2pc doesn't reflect attentional capture per se, but indexes downstream processes associated with identification of all objects within an attended area.

Our results show that attention capture in EIB is not automatic despite the robust nature of attention capture associated with this paradigm. Attention capture in the standard EIB paradigm is probably robust for several reasons.

First, participants are actively attending to the picture stream in order to detect the target picture. In our Experiment 2 and in the MOT-Only condition of Experiment 1, they are fully attending to the moving objects in order to track them and attempting to ignore the picture stream. Second, in EIB, the target picture is essentially a "singleton" (an object with a unique feature relevant to all the other objects; e.g., a red object in a set of green objects) because the "horizon line", which is a strong feature of the target and background pictures (see Fig. 12), is oriented vertically for the target while it is horizontal for the remaining background pictures. Therefore, participants may adopt a strategy of detecting "singletons" (Bacon & Egeth, 1994). In other words, they are set for a picture that is different than the rest of the pictures in the stream. If so, they wouldn't be able to immediately suppress the distractor picture when it appears because it is also a singleton (it is the only picture with close-ups of people and animals). Instead, suppression could only occur after the nature of the picture is determined. That would explain why, in the present experiments, there is a P_D component elicited by task-irrelevant distractors that is early enough to affect the EPN component. In the standard EIB paradigm, the early P_D is absent, reflecting delayed suppression (Kennedy et al., 2014). In the current studies, the gap appears as a change to an existing object which reduces onset transients associated with the appearance of a new object. Presumably the signal accompanying this change is quite different than the onset transient associated with the appearance of the distractor picture allowing it to be rapidly suppressed. Rapid suppression of the distractor would terminate processing before the emotional information is retrieved reducing the salience of emotional pictures. In EIB, suppression may sometimes be late enough to allow emotional information to emerge and capture attention. Most other paradigms used to study emotional capture do not have this confound of physical and emotional salience and will often elicit smaller or absent capture effects.

5. Summary

EIB is associated with particularly robust emotional capture effects and some researchers (e.g., McHugo et al., 2013) have speculated that it may not be affected by top-down control processes. In an initial study, we found that late stages in the EIB, reflected in the P3b component are strongly dependent on attention. However earlier components, such as the EPN, were not affected by the amount of attention allocated to the picture stream. And surprisingly, MOT performance was not affected by whether distractors were negative or neutral in valence. In a follow-up study we showed that distractors produced a brief capture event that was the same magnitude for negative and neutral distractors. This indicates that initial capture of attention in EIB is driven by physical salience of the distractors relative to the background stream pictures. In other research, we showed that when the physical salience of the emotional pictures is reduced by embedding them in similar, neutral background pictures, the EIB effect is reduced by approximately 75%. These results show that the difference between negative and neutral distractor pictures in the EIB paradigm can be eliminated by engaging observers in an attention demanding task and instructing them to ignore the picture stream. It also appears that an important contributing factor to the robust emotional capture effects in EIB is the use of background pictures which are physically different than the distractors allowing capture to be driven by the physical salience of the distractors.

Author note

This research was supported by NSF Grant BCS-1059560 to JEH and NSF EPSCoR grant 1632849 to the University of Delaware. Correspondence relating to this article may be sent to James E. Hoffman, Department of Psychological and Brain Sciences, University of Delaware, Newark, DE 19716 (hoffman@psych.udel.edu).

REFERENCES

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., Dimase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance, 31*(4), 643–667. https://doi.org/10.1037/0096-1523.31.4.643a.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences, 16*(8), 437–443.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics, 55*(5), 485–496.

Baker, A. L., Hoffman, J. E., & Turco, T. (2018). The role of physical salience capture in emotion-induced blindness (EIB). In *Poster presented at the 26th annual meeting of the object perception, attention, and memory workshop, New Orleans, LA*.

Cardinal, R. N., & Aitken, M. R. (2013). *ANOVA for the behavioral sciences researcher*. Psychology Press.

Choisdealbha, Á. N., Piech, R. M., Fuller, J. K., & Zald, D. H. (2017). Reaching back: The relative strength of the retroactive emotional attentional blink. *Scientific Reports, 7*, 43645.

Cohen, M. C., Alvarez, G. A., & Nakayama, K. (2011). Natural scene perception requires attention. *Psychological Science, 22*(9), 1165–1172.

D'Andrea-Pennaa, G. M., Frank, S. M., Heatherton, T. F., & Tse, P. U. (2017). Distracting tracking: Interactions between negative emotion and attentional load in multiple-object tracking. *Emotion, 17*(6), 900–904.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Donchin, E., & Coles, M. G. (1998). Context updating and the P300. *Behavioral and Brain Sciences, 21*(1), 152–154.

Eimer, M. (2015). EPS Mid-Career Award 2014: The control of attention in visual search: Cognitive and neural mechanisms. *The Quarterly Journal of Experimental Psychology, 68*(12), 2437–2463.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology, 18*(4), 1030–1044.

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: Evidence from temporal-spatial PCA. *Psychophysiology, 46*(3), 521–530.

Gaspelin, N., Leonard, C. J., & Luck, S. J. (2017). Suppression of overt attentional capture by salient-but-irrelevant color singletons. *Attention, Perception, & Psychophysics, 79*(1), 45–62.

Gaspelin, N., & Luck, S. J. (2019). Inhibition as a potential resolution to the attentional capture debate. *Current Opinion in Psychology, 29*, 12–18.
Hajcak, G., Weinberg, A., MacNamara, A., & Foti, D. (2011). ERPs and the study of emotion. In S. J. Luck, & E. S. Kappenman (Eds.), *Oxford handbook of ERP components (441-474)*. New York: Oxford University Press.
Hickey, C., Di Lollo, V., & McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *Journal of Cognitive Neuroscience, 21*(4), 760–775.
Holmes, A., Kiss, M., & Eimer, M. (2006). Attention modulates the processing of emotional expression triggered by foveal faces. *Neuroscience Letters, 394*(1), 48–52.
Holmes, A., Nielsen, M. K., Tipper, S., & Green, S. (2009). An electrophysiological investigation into the automaticity of emotional face processing in high versus low trait anxious individuals. *Cognitive, Affective, & Behavioural Neuroscience, 9*(3), 323–334.
JASP Team. (2018). *JASP (version 0.9) [Computer software]*.
Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.
Kennedy, B. L., Rawding, J., Most, S. B., & Hoffman, J. E. (2014). Emotion induced blindness reflects competition at early and late processing stages: An ERP study. *Cognitive, Affective, & Behavioral Neuroscience,* 14(4), 1485–1498. https://doi.org/10.3758/s13415-014-0303-x.
Kim, M., & Hoffman, J.E. (in preparation). Emotional pictures elicit an EPN component in the absence of attention capture. Paper in preparation.
Kim, M., Taylor, M., & Hoffman, J. E. (2017). Emotional pictures automatically capture attention. *Journal of Vision, 17*(10), 1292-1292.
Kiss, M., Driver, J., & Eimer, M. (2009). Reward priority of visual target singletons modulates event-related potential signatures of attentional selection. *Psychological Science, 20*(2), 245–251.
Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology, 45*(2), 240–249.
Kranczioch, C., Debener, S., Maye, A., & Engel, A. K. (2007). Temporal dynamics of access to consciousness in the attentional blink. *Neuroimage, 37*(3), 947–955.
Lachaud, C. M., & Renaud, O. (2011). A tutorial for analyzing human reaction times: How to filter data, manage missing values, and choose a statistical model. *Applied Psycholinguistics, 32*(2), 389–416.
Landman, R., Sharma, J., Sur, M., & Desimone, R. (2014). Effect of distracting faces on visual selective attention in the monkey. *Proceedings of the National Academy of Sciences, 111*(50), 18037–18042.
Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Technical Report A-8. Gainesville, FL: University of Florida.
Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 451–468.
Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171.
Luck, S. J. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. *The Oxford Handbook of Event-Related Potential Components*, 329–360.
Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146–157.
MacLeod, J., Stewart, B. M., Newman, A. J., & Arnell, K. M. (2017). Do emotion-induced blindness and the attentional blink share underlying mechanisms? An event-related potential study of emotionally-arousing words. *Cognitive, Affective, & Behavioral Neuroscience, 17*(3), 592–611.
McHugo, M., Olatunji, B. O., & Zald, D. H. (2013). The emotional attentional blink: What we know so far. *Frontiers in Human Neuroscience, 7*, 151.
Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y. H., et al. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience, 19*(8), 1041–1049.
Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*(1), 56–60.
Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review, 12*(4), 654–661.
Most, S. B., Smith, S. D., Cooter, A. B., Levy, B. N., & Zald, D. H. (2007). The naked truth: Positive, arousing distractors impair rapid target perception. *Cognition & Emotion, 21*, 964–981.
Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods, 192*(1), 152–162.
Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466.
Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*(11), 773–783.
Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences, 99*(17), 11458–11463.
Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biological Psychology, 92*(3), 492–512.
Rafal, R. D., Koller, K., Bultitude, J. H., Mullins, P., Ward, R., Mitchell, A. S., et al. (2015). Connectivity between the superior colliculus and the amygdala in humans and macaque monkeys: Virtual dissection with probabilistic DTI tractography. *Journal of Neurophysiology, 114*(3), 1947–1962.
Rellecke, J., Sommer, W., & Schacht, A. (2012). Does processing of emotional facial expressions depend on intention? Time-resolved evidence from event-related brain potentials. *Biological Psychology, 90*(1), 23–32.
Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374.
Sawaki, R., Geng, J. J., & Luck, S. J. (2012). A common neural mechanism for preventing and terminating the allocation of attention. *The Journal of Neuroscience, 32*(31), 10725–10736.
Sawaki, R., & Luck, S. J. (2010). Capture versus suppression of attention by salient singletons: Electrophysiological evidence for an automatic attend-to-me signal. *Attention, Perception, & Psychophysics, 72*(6), 1455–1470.
Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: Event-related brain potential studies. *Progress in Brain Research, 156*, 31–51.
Schupp, H. T., Stockburger, J., Bublatzky, F., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2007). Explicit attention interferes with selective emotion processing in human extrastriate cortex. *Bmc Neuroscience, 8*(1), 16.
Schupp, H. T., Stockburger, J., Codispoti, M., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2007). Selective visual attention to emotion. *Journal of Neuroscience, 27*(5), 1082–1089.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience, 8*(10), 1391–1400.
Shapiro, K., Schmitz, F., Martens, S., Hommel, B., & Schnitzler, A. (2006). Resource sharing in the attentional blink. *Neuroreport, 17*(2), 163–166.
Sibly, M. (2005). Blitz3D *(version 1.90) [computer software]*. Auckland, New Zealand: Blitz Research Ltd.
Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51*(6), 599–606.
Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135*(2), 77–99.
Tran, A., & Hoffman, J. E. (2016). Visual attention is required for multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance, 42*(12), 2103–2114. https://doi.org/10.1037/xhp0000262.
Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for postperceptual suppression during the attentional blink. *Journal of Experimental Psychology. Human Perception and Performance, 24*(6), 1656–1674.
Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron, 30*(3), 829–841.
Wang, P., & Nikolic, D. (2011). An LCD monitor with sufficiently precise timing for research in vision. *Frontiers in Human Neuroscience, 5*, 85.
Wiens, S., & Syrjänen, E. (2013). Directed attention reduces processing of emotional distracters irrespective of valence and arousal level. *Biological Psychology, 94*(1), 44–54.
Yantis, S., & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance, 20*(1), 95–107.
Zivony, A., Allon, A. S., Luria, R., & Lamy, D. (2018). Dissociating between the N2pc and attentional shifting: An attentional blink study. *Neuropsychologia, 121*, 153–163.