Emotion
2011, Vol. 11, No. 5, 1134–1143

1528-3542/11/$12.00 DOI: 10.1037/a0022586

No Emotional "Pop-Out" Effect in Natural Scene Viewing

David J. Acunzo
University of Edinburgh

John M. Henderson
University of South Carolina

It has been shown that attention is drawn toward emotional stimuli. In particular, eye movement research suggests that gaze is attracted toward emotional stimuli in an unconscious, automated manner. We addressed whether this effect remains when emotional targets are embedded within complex real-world scenes. Eye movements were recorded while participants memorized natural images. Each image contained an item that was either neutral, such as a bag, or emotional, such as a snake or a couple hugging. We found no latency difference for the first target fixation between the emotional and neutral conditions, suggesting no extrafoveal "pop-out" effect of emotional targets. However, once detected, emotional targets held attention for a longer time than neutral targets. The failure of emotional items to attract attention seems to contradict previous eye-movement research using emotional stimuli. However, our results are consistent with studies examining semantic drive of overt attention in natural scenes. Interpretations of the results in terms of perceptual and attentional load are provided.

Keywords: visual attention, eye movement, emotional processing, scene perception

Visual attention is a key mechanism of human cognition, enabling us to select relevant visual stimuli by prioritizing the processing of certain features or aspects of the incoming information. Characterizing attentional processes is, therefore, a *sine qua non* for the understanding of cognition. A crucial issue is the extent to which high-level information, such as semantic or emotional information, plays a role in the exogenous drive of covert and overt attention.

Emotional stimuli, by definition, are stimuli with high motivational value, important for survival of the individual or the species. From an evolutionary point of view, individuals who can detect and react to these stimuli fast will be advantaged. The capture of attention by emotional stimuli has, therefore, received a great deal of interest. It is now known that these stimuli can capture and hold attention more easily than neutral stimuli, but the neural mechanisms of these interactions are yet to be understood (Vuilleumier, 2005).

Visual search paradigms have been used to assess whether emotional and, in particular, fear-relevant targets are detected faster than fear-irrelevant targets. Many of these studies have found a search advantage for emotional items, such as snakes, in comparison to neutral items, such as mushrooms (e.g., Öhman, Flykt, & Esteves, 2001; Flykt, 2005; Blanchette, 2006; Fox, Griggs, & Mouchlianitis, 2007). However, further studies have shown that the situation is more complex (see, e.g., Tipples, Young, Quinlan, Broks, & Ellis, 2002; Lipp, Derakshan, Waters, & Logies, 2004; Soares, Esteves, & Flykt, 2009; Flykt, 2006). Cave and Batty (2006) interpreted these contrasting results as follows: "[T]hreat itself, as opposed to [visual] features associated with threat, seems to be less of a factor in visual search than was first suggested" (p. 636). In line with this hypothesis, Coelho, Cloete, and Wallis (2010) used schematic face and face-like stimuli and suggested that the search advantage for particular facial expression is driven by low-level features. As noted by Soares et al. (2009), discrepancies between experimental results are probably caused by variations in the search tasks, making it difficult to draw definite conclusions about the search mechanisms.

Eye-movement research has also shown that emotional stimuli attract overt attention (Calvo & Lang, 2004; Nummenmaa, Hyönä, & Calvo, 2006, 2009; Alpers, 2008). In these studies, participants were peripherally presented two images while their eye movements were recorded. When an emotional image was presented concurrently with a neutral image, the probability of the first fixation landing on the emotional picture was higher. It was also shown that participants fixated the emotional image for a longer time. In Nummenmaa et al. (2006), even when explicitly instructed to attend to the neutral image, participants first fixated on the emotional image. Finally, a more recent study (Nummenmaa et al., 2009) examined saccade latency when participants were instructed to look either left or right when a distractor image was presented on each side. It was found that saccade latency was delayed when the image opposite to the instructed direction was emotional. Saccade trajectories were also modulated by surrounding emo-

This article was published Online First July 25, 2011.

David J. Acunzo, Neuroinformatics Doctoral Training Center, Institute for Adaptive & Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, United Kingdom; John M. Henderson, Department of Psychology, University of South Carolina.

The authors thank Dr. Tim J. Smith for his helpful comments on the manuscript of this article. This research was supported by an Economic and Social Research Council (ESRC) grant (to J.M.H.) and by the Neuroinformatics Doctoral Training Centre, funded by the Life Sciences Interface programme of the Engineering and Physical Sciences Research Council (EPSRC), and cofunded by the Medical Research Council (MRC) and the Biotechnology and Biological Sciences Research Council (BBSRC).

Correspondence concerning this article should be addressed to David J. Acunzo, Neuroinformatics DTC, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom. E-mail: david.acunzo@ed.ac.uk

tional content: When participants were instructed to saccade vertically while presented distractor images on the sides, the saccade curved away from emotional images. In Becker and Detweiler-Bedell (2009), participants were instructed to passively look at an array of four faces while their eye movements were recorded. Interestingly, the researchers found that participants avoided looking at the threatening face as early as the first saccade, suggesting an early evaluation of the face valence, biasing subsequent eye movements.

These results suggest that eye movements are modulated by emotional content within the visual field in an unconscious, automated manner. This is consistent with paradigms looking at covert attention, which suggest that emotional stimuli can modulate attention even when they are task-irrelevant (Bar-Haim, Lamy, Pergamin, Backermans-Kranenburg, & van Ijzendoorn, 2007).

However, the research discussed above used particular viewing conditions. First, the stimuli used usually contained a small number of independent images or items. Because the items were independent (i.e., content and location were unrelated), an independent "emotional gist" could have been extracted for each item. Additionally, the items were often presented extrafoveally while the participant was fixating on a dot in the center of the screen. This low initial foveal load might have facilitated the emotional processing of the images. Finally, the high frequency of emotional stimuli, together with the low variance of semantic content, and the few possible locations where items could be displayed may have eased the task of the participants by increasing the expectation of the participant for emotional stimuli. If previous research shows that attention and eye movements are modulated by extrafoveal emotional content under these particular conditions, it is unclear whether these effects would remain under more natural conditions where perceptual and foveal load is high and where objects are part of a whole scene.

To answer this question, we adapted a paradigm initially developed to assess the effects of semantic gist violation on eye movements (Loftus & Mackworth, 1978). Participants' eye movements were recorded while viewing scenes in which one target object did not fit with the rest of the image (e.g., an octopus in a farm). Interestingly, items violating the gist do not seem to generate any semantic "pop-out" effect, but do hold attention longer than nonviolating items (De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999; Gareze & Findlay, 2007; Castelhano, Mack, & Henderson, 2009; Võ & Henderson, 2009; Rayner, Castelhano, & Yang, 2009; but see Loftus & Mackworth, 1978; Becker, Pashler, & Lubin, 2007; Underwood & Foulsham, 2006; Gareze & Findlay, 2007; Underwood, Templeman, Lamming, & Foulsham, 2008).

We developed a set of stimuli consisting of pairs of realistic scenes. Each pair consisted of two photographs, which solely differed by a target item: In one condition, this item had a neutral valence, and in the other, it had an emotional (i.e., positive or negative) valence. Participants were asked to try to remember those images for a subsequent memory test while their eye movements were recorded. Additionally, scenes were horizontally flipped, so that the target item was presented on the left or right side of the initial fixation point. Target item position (left or right) and valence (neutral or emotional) were then the conditions of our 2×2 within-participant design.

According to previous research on eye movements and emotional stimuli, we hypothesized that emotional items would pop-out and be fixated earlier than neutral targets. However, research on eye-movement and scene perception suggests that no such pop-out should occur. Further, we tested the hypothesis that emotional targets would be fixated earlier if located on the left-hand side of the initial fixation point. Previous research suggests laterality effects, with a right hemisphere advantage to process emotional stimuli (see, e.g., Keil, Morati, Sabatinelli, Bradley, & Lang, 2005; Calvo & Nummenmaa, 2007; Calvo & Avero, 2008). In particular, in the context of eye-movement research, Alpers (2008) used the same paradigm as Calvo and Lang (2004) and reported that the effects of the emotional content on the first fixation observed in previous experiments were present only when the emotional picture was presented in the left hemifield. Finally, we hypothesized that participants would fixate on the emotional targets for a longer time than the neutral ones, as both the literature on emotional stimuli (Calvo & Lang, 2004, and replications) and scene perception (De Graef et al., 1990; Henderson et al., 1999; Gareze & Findlay, 2007; Castelhano et al., 2009; Võ & Henderson, 2009; Rayner et al., 2009) would suggest.

Method

The procedure and stimuli were approved by the University of Edinburgh Department of Psychology Ethics Committee.

Participants

Sixteen participants (10 female) participated in the experiment, most of whom were students in the University of Edinburgh recruited through an internal university website. All participants reported normal or corrected-to-normal vision. They were compensated £6/hr.

Stimulus Material

Stimulus design. The stimulus material consisted in 48 full-color 24 bit images of maximal resolution of 800×600 pixels. Each scene conformed to one of the 2×2 conditions: emotional versus neutral and left versus right. In the emotional condition, a target item in the scene was emotionally evocative; in the neutral condition, an emotionally neutral target replaced the emotionally evocative target. In the left and right conditions, the target was located in the left and right part of the image, respectively, generated by simply mirroring the entire image over a vertical axis. Half of the emotional stimuli contained positive targets and the other half contained negative ones. Examples of negative targets included people with facial tumors, a threatening animal (e.g., snake), a face showing fear, and a face covered with blood. Positive targets represented people hugging or kissing, children playing, or fluffy animals. Neutral targets included bags, faces, or whole characters. An example of a scene is shown in Figure 1.

Emotional valence and arousal were controlled by asking a population of 16 participants (10 female) who did not take part in the main experiment to rate the target items. Emotional targets were rated significantly higher in terms of arousal than neutral stimuli. Positive (negative) targets were rated significantly higher

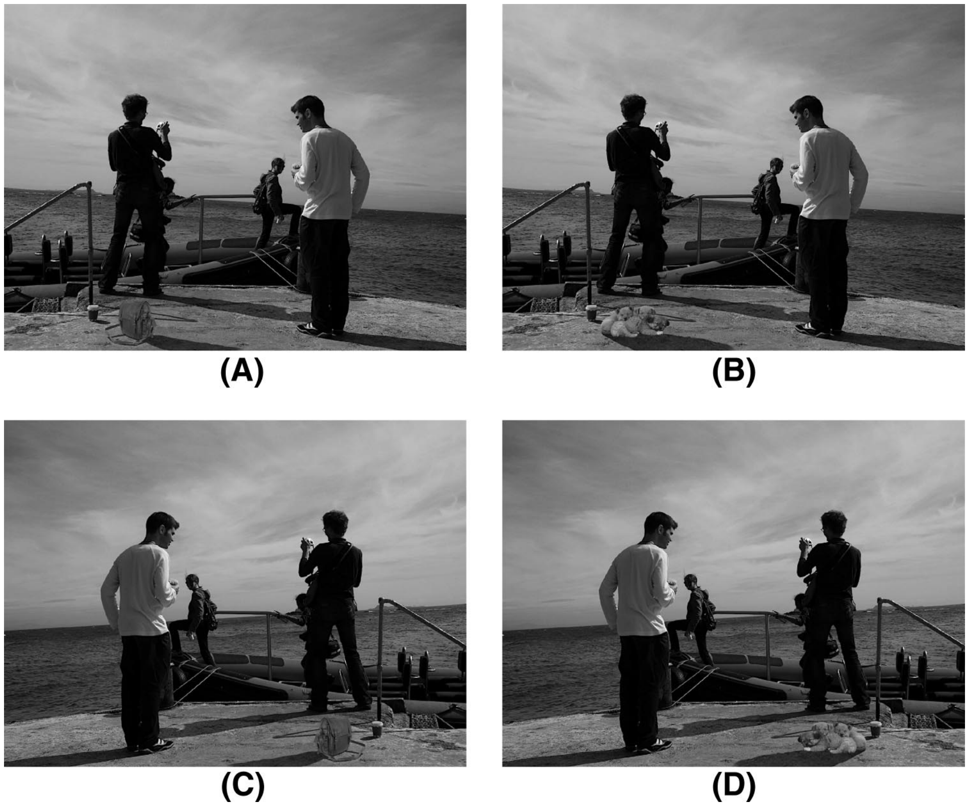

Figure 1. Example of scene presented to the participants. The four conditions of an example scene: neutral-left (A), emotional-left (B), neutral-right (C), and emotional-right (D). The neutral and emotional conditions differ by one item (bag/dogs), whereas the left and right conditions are the horizontally flipped versions of one another. Each participant was presented only one of the four conditions of every scene. Scenes were presented in full color.

(lower, respectively) in terms of valence than their neutral counterpart. More details on the procedure and results of the stimulus validation study are given in the Appendix.

For each pair of dual images (i.e., pair of same background images with a different target), we defined a common target interest area (IA) that included the neutral and emotional target for both images. The IAs of the mirrored images had mirrored IAs from the original image.

In the neutral condition, 9 of the 12 scenes were artificially modified, and all 12 were modified in the emotional condition. A modification involved either the addition or alteration of the target item. Scenes were found on the Internet (except one, which was a photograph taken by a member of the research group), whereas targets were taken from the Internet, the International Affective Picture System (IAPS; Lang, Bradley, & Curthbert, 2008), the NimStim face database (Tottenham et al., 2009), and Hemera Photo-Objects 2.07 (Hemera Technologies, Seattle, WA). Images were manipulated using GIMP 2.4.0 (available at: http://www.gimp.org; accessed January 24, 2011). Targets were adapted for luminance, saturation, color, and contrast in order to make the addition or replacement as natural as possible. In many cases, manual modification of the lighting of the target was necessary, and shadows and reflections were modified or added for more realism. To ensure that the modifications did not lead to a difference of saliency between the neutral and emotional conditions, we ran the Matlab implementation of a saliency model (Itti & Koch, 2000) on our images. The saliency map was computed and normalized for the images. The average saliency was then computed within the target IA. A Wilcoxon signed-rank test was used to compare the average saliency within the 12 IAs containing a neutral target (Mdn = 0.0247), with the 12 containing an emotional target (Mdn = 0.0289). Differences were not significant: T = 31, $p > .8$, $r = .04$.

Apparatus

Images were presented on a 21″ CRT monitor at a viewing distance of 90 cm with a refresh rate of 140 Hz. Their maximum resolution was 800 × 600 pixels, subtending a maximum visual angle of 25.7 × 19.4 degrees. Eye movements were monitored by an SR Research EyeLink 1000 eye-tracker (SR Research, Ltd., Kanata, Ontario, Canada). The head of the participant was fixed on a chin-rest. Fixation position was sampled at 1,000 Hz, and saccades prior to critical fixations were detected using a 17-sample saccade detection model with a velocity threshold of 30 deg/s, an acceleration threshold of 8,000 deg/s^2, and a minimum amplitude of 0.5 degrees. The right eye only was tracked, whereas viewing was binocular. The experiment was controlled with SR Research Experiment Builder software.

Procedure

Each participant was presented a consent form to be signed, informing about the experiment and the emotionally evocative nature of some of the stimuli. Before the viewing task, the participants were given the State Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970) questionnaire to fill in.

Each participant was presented one of four blocks, each containing all 12 scenes. Each block contained three scenes in each of the four conditions: neutral-left (NL), neutral-right (NR), emotional-left (EL), and emotional-right (ER). No participant was presented the same scene in more than one condition. The order of image presentation was randomized within the assigned block for each participant.

Participants were told that they would be shown 12 images for 15 s each, and that they would have to memorize them for a subsequent memory task. The memory task was never given. Calibration of the eye tracker, using nine points on the screen, was performed, followed by a validation. At the beginning of each trial, a point in the middle of the screen had to be fixated by the participant, for a fixation check. The trial was then initiated manually by the experimenter. If inaccuracy of the eye-tracker was detected, a new calibration was performed.

Eye-Movement Data Manipulation

Raw data were first filtered and preprocessed with SR Data Viewer. Most data manipulation was carried out using Matlab 7.0 (MathWorks, Inc., Natick, MA). Graphs and statistical tests were done with Matlab and SPSS (SPSS, Inc., Chicago, IL). Analyses of variance (ANOVAs) included valence (neutral, emotional) and side (left, right) as within-participant factors. When possible, *t* tests were performed to compare two independent samples. Otherwise, a Wilcoxon's rank-sum test was conducted.

No fixation within the IA occurred during the scene presentation for 16 trials (4 NL, 2 NR, 4 EL, and 6 ER; 8.33% of all trials). Those trials were not included in the following analyses.

Results

Capture of Attention

One of the main questions that we address here is whether emotional targets attract attention more than neutral ones when embedded in a natural scene. Table 1 and Figure 2 summarize the statistics reported in this section.

Number of fixations and latency to IA. To assess the difference in attentional capture by targets across conditions, we examined how early in scene exploration the targets were fixated. To do so, we looked at the difference in the number of fixations prior to the first fixation within the IA, and at the amount of time spent exploring the scene prior to the first fixation within the IA. For the number of fixations to IA, no effect of valence $F(1, 15) < 1$ ($\omega^2 = .032$) or side $F(1, 15) < 1$, ($\omega^2 = .025$) was found. A nonsignificant valence-side interaction $F(1, 15) = 3.74$, $p > .05$ ($\omega^2 = .447$) was found, with fewer fixations to IA for the neutral-left and emotional-right conditions, compared with neutral-right and emotional-left, whereas we would have expected fewer fixations for the emotional-left versus emotional-right.

Latency to IA showed the same pattern: no effect of valence $F(1, 15) < 1$ ($\omega^2 = .027$), side $F(1, 15) < 1$, ($\omega^2 = .022$) and a nonsignificant interaction $F(1, 15) = 3.87$, $p > .05$ ($\omega^2 = .453$).

Incoming saccade amplitude to IA. The amplitude of the first saccade ending within the IA provides information about extrafoveal processing of emotional targets. Given the hypothesis that emotional targets capture attention extrafoveally, we should observe a larger saccade amplitude for the emotional condition than for the neutral one. The analyses showed a nonsignificant trend, with longer saccade amplitude to emotional targets $F(1, 15) = 2.83$, $p > .05$ ($\omega^2 = .399$), no effect of side $F(1, 15) = 1.45$, $p > .05$ ($\omega^2 = .297$), and no interaction, $F(1, 15) < 1$ ($\omega^2 = .052$).

Hold of Attention

In this subsection, we analyzed events occurring once the target was overtly attended and compared the hold of attention by the emotional targets against the neutral ones. After the target was fixated, its location within the image was not a relevant variable in these analyses. We consequently collapsed the left and right conditions for this part of the analysis. The remaining conditions were simply emotional and neutral. Table 2 and Figure 3 summarize the statistics reported in this section.

First fixation duration within IA. The durations of the first fixation on the target item can be indicative of the encoding of the fixated object (see, e.g., Henderson & Hollingworth, 1999), although recent evidence suggest that fixation durations are only partially driven by visual input (Henderson & Smith, 2009; Nuthmann, Smith, Engbert, & Henderson, 2010). The durations of the first fixation within the IA for neutral (*Mdn* = 255.3 ms) and emotional (*Mdn* = 249.4 ms) targets showed no significant difference with Wilcoxon's rank-sum test: T = 282, $p > .5$, $r = .16$.

Table 1
Capture of Attention: Summary

	Conditions				*F* values		
	NL	NR	EL	ER	Em	S	Em × S
Number of fixations to IA	6.49 (.76)	8.28 (1.33)	8.36 (1.47)	6.78 (.82)	<1	<1	3.74
Latency to IA (ms)	1,525 (206.4)	2,069 (396.2)	2,090 (435.8)	1,603 (259.7)	<1	<1	3.87
Saccade amp to IA (deg)	5.7 (.35)	6.0 (.44)	6.9 (0.54)	7.2 (.60)	2.83	<1	<1

Note. Mean and (standard error) of the "capture of attention" variables for each of the four conditions: neutral-left (NL), neutral-right (NR), emotional-left (EL), and emotional-right (ER). *F*-ratio obtained from the repeated measures, with the factors emotion (Em) and side (S). None of the *F* values obtained were statistically significant.

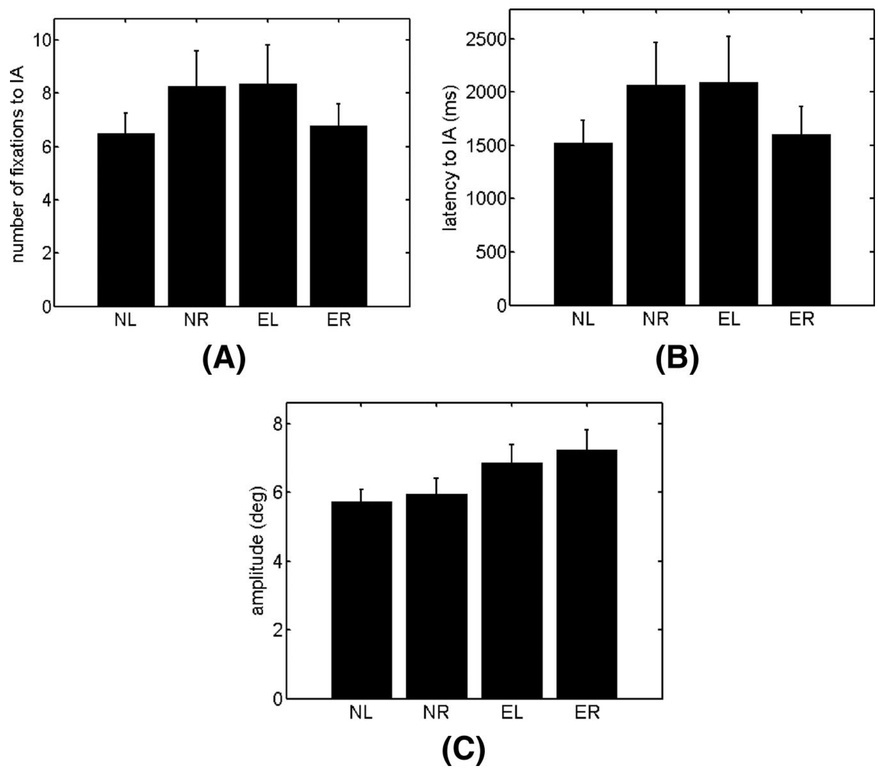

Figure 2. Capture of attention. Mean value and standard error of the number of fixations (A), latency (B) to IA, and amplitude of the first fixation to IA (C) across the four conditions. NL and NR indicate neutral left and neutral right, respectively, and EL and ER indicate emotional left and emotional right. See numeric values in Table 1.

First-pass number of fixations and time. We measured the number of fixations and time spent between the very first fixation within the target IA and the first subsequent fixation outside the IA.

First-pass number of fixations showed a significant difference between the neutral (M = 2.40, SE = 0.21) and emotional (M = 3.92, SE = 0.37) conditions: $t(23.3) = 3.56$, $p < .02$. More fixations were placed on the emotional than neutral target.

Similarly, the time spent during the first pass of the IA was longer for the emotional targets (M = 1253.7 ms, SE = 93.9) than for the neutral targets (M = 666.1 ms, SE = 71.4): $t(30) = 4.98$, $p < 10^{-4}$.

Total number of fixations and dwell time. Finally, we measured the total number of fixations and the total fixation time spent within the IAs.

The average total number of fixations within the IA for emotional targets (M = 11.40, SE = 1.52) was not significantly higher than for neutral targets (M = 8.22, SE = 0.83): $t(23.14) = 1.84$, $p > .07$. However, the average total time spent within the IA was significantly longer for the emotional targets (M = 3748.4 ms,

Table 2
Hold of Attention: Summary

	Conditions		Comparison	
	Neutral	Emotional	Stat.	Value
First fixation duration (ms)	*Mdn* 255.3	*Mdn* 249.4	T	282
First-pass number of fixations	2.40 (.21)	3.92 (.37)	t	3.56*
First-pass duration (ms)	666.1 (71.4)	1,253.7 (93.9)	t	4.98**
Total number of fixations in IA	8.22 (.83)	11.40 (1.52)	t	1.84
Total IA dwell time (ms)	2,464.2 (241.4)	3,748.4 (444.8)	t	2.54*

Note. Median or mean and (standard error) of the "hold of attention" variables for the two conditions: neutral and emotional. Rank-sum (T) or t-score (t) values from the two independent sample comparison are given.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

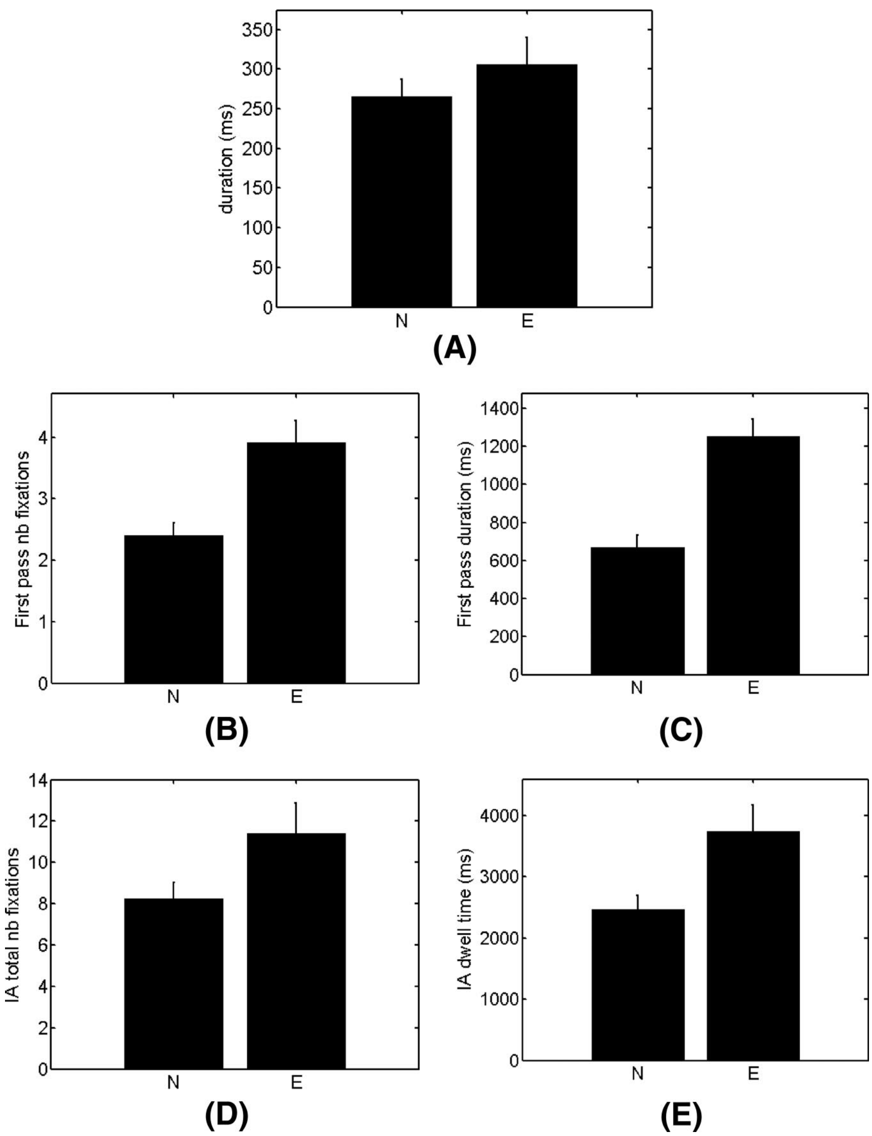

Figure 3. Hold of attention. Mean value and standard error of the first IA fixation duration (A), the first-pass number of fixations (B) and duration (C), the total IA number of fixations (D), and dwell time (E). N and E indicate neutral and emotional conditions, respectively. See numeric values in Table 2.

SE = 444.8) than for the neutral targets (M = 2464.2 ms, SE = 241.4): $t(23.1) = 2.54$, $p < .02$.

Discussion

One of the main aims of this study was to test for earlier detection of emotional target items when embedded within an entire natural image. Previous research suggests the existence of an exogenous drive of eye movements by peripherally attended emotional stimuli (see Calvo & Lang, 2004; Nummenmaa et al., 2006, 2009; Alpers, 2008; Becker & Detweiler-Bedell, 2009). However, our experiment suggests that when embedded in a scene, this exogenous drive disappears. At the same time, we found that once fixated, emotional items hold attention longer than neutral ones, which is in line with previous research reporting delayed attention disengagement or hold of attention to emotional stimuli (Calvo & Lang, 2004; Nummenmaa et al., 2006; Fox, Russo, Bowles, & Dutton, 2001; Fox, Russo, & Dutton, 2002) and part of the scene perception research reporting hold of attention to semantically incongruous items (De Graef et al., 1990; Henderson et al., 1999; Gareze & Findlay, 2007; Castelhano et al., 2009; Võ & Henderson, 2009; Rayner et al., 2009).

We also found no effect of the position of the target stimulus, whereas previous research has suggested a right-hemisphere advantage for emotional stimuli processing (e.g., Keil et al., 2005; Calvo & Nummenmaa, 2007; Calvo & Avero, 2008). In particular, Alpers 2008) used a paradigm similar to Calvo and Lang (2004) and found that emotional stimuli were fixated earlier only when positioned on the left visual hemifield. It should be noted, however, that tighter controls in Alpers (2008) would have made the claim stronger (e.g., comparison of saliency between stimulus groups; an additional condition to assess a potential eye-movement bias toward the left side).

The lack of attentional capture by emotional stimuli cannot be attributed to a lack of differential emotional impact on the participants, because significant differences in valence and arousal were found from the values given by independent raters (see Appendix). We also controlled for low-level saliency, using a computational saliency model (Itti & Koch, 2001): No difference between the two target groups was found.

We tried to cover a wide range of stimuli (e.g., fearful face, couple kissing, animal, etc.). This heterogeneity in the nature of the target might be seen as a weakness, because different emotions are not processed the same way and will not generate the same reaction: A cute cat will probably not attract the eye the same way as a face covered with blood. Further, given the small number of stimuli, we did not look separately at the differential effects of positively and negatively valenced targets. This said, previous studies on emotion have also used heterogeneous emotional stimuli, in particular when using the IAPS database, and the effects of negatively and positively valenced stimuli on eye movements reported, so far, are qualitatively similar, with perhaps faster attentional capture by negative stimuli and longer hold of attention by positive stimuli (Calvo & Lang, 2004; Nummenmaa et al., 2006; Alpers, 2008). Importantly, we found significantly higher IA first-pass number of fixations and dwell time for emotional stimuli, which supports the idea that our manipulation was strong enough to elicit a modulation in the scene-exploration process.

Although our results are inconsistent with previous eye-movement studies looking at emotional stimuli in isolation, they are highly consistent with the body of data looking at eye movements and scene perception (De Graef et al., 1990; Henderson et al., 1999; Gareze & Findlay, 2007; Castelhano et al., 2009; Võ & Henderson, 2009). The majority of experimental results indicate that gist-inconsistent targets do not elicit earlier fixations than gist-consistent ones. However, they do hold attention longer once fixated. We can argue that both gist-inconsistent and emotional items are behaviorally relevant. Gist-inconsistent items are more informative about the environment than gist-consistent ones, whereas emotional items are behaviorally relevant because of their intrinsic motivational value. This is illustrated by the fact that both gist-inconsistent and emotional items are fixated more than gist-consistent and neutral ones, respectively. It should be noted, though, that earlier fixations to inconsistent objects have been reported in some studies (Loftus & Mackworth, 1978; Becker et al., 2007; Underwood & Foulsham, 2006; Gareze & Findlay, 2007; Underwood et al., 2008). Interestingly, this discrepancy between studies has been partly attributed to a difference in sparsity of the scenes (Võ & Henderson, 2009). Less-cluttered scenes enable participants to detect semantic inconsistencies more easily.

Similarly, differences in experimental design and stimuli are likely to account for the differences between our results and previous eye-movement studies using emotional stimuli. First, each stimulus used in our study consisted of an individual scene presented on a full screen. This is in contrast with the paradigms used in Calvo and Lang (2004), Nummenmaa et al. (2006, 2009), and Becker and Detweiler-Bedell (2009), which consisted of two or four peripherally presented images with a fixation point in the center of the screen. In our case, foveal load was high from stimulus onset, which was not the case in the other paradigms. Our results are in line with Calvo and Nummenmaa (2007), who reported that foveal load impairs the processing of peripherally presented emotional stimuli.

Second, our target items were embedded in a whole image. Target search and previous eye-tracking paradigms have focused on the effects of images presented simultaneously to the participant. In those studies, the images are probably seen by the participant as independent, unrelated entities, which are localized and separated in the visual field and can contain unrelated objects. Each of them can thus be processed as a whole, independently from each other, and an emotional and semantic gist can be extracted from each entity. In our case, objects cannot be seen independently, because they are all linked within the image. Additionally, some of our emotional target items were significantly smaller than the images used in previous research (and, in particular, Calvo & Lang, 2004; Nummenmaa et al., 2006, 2009). In Calvo, Nummenmaa, and Hyönä (2008), it is suggested that the processing of the emotional gist of images may come from a "fast" subcortical route (see Le Doux, 1995), which would project to the amygdala, via the superior colliculus. Neurons of the superior colliculus respond to stimuli situated peripherally and containing low spatial frequencies (Miller, Pasik, & Pasik, 1980; Rodman, Gross, & Albright, 1989); therefore, some of our small-sized targets might not have been able to activate this pathway. However, the involvement of this subcortical pathway in visual emotional processing is still debated (see Pessoa & Ungerleider, 2004; Storbeck, Robinson, & McCourt, 2006). In any case, we can hypothesize that the effects observed in previous studies are due to the "emotional gist" of individual images. If this is the case, our results make sense, because the target items were not seen as independent from the rest of the scene. Semantic and emotional information for each element of the image was thus more difficult to process.

Third, the explicit task given to the participant was to memorize the scenes for a subsequent memory test. This task was unrelated to emotional appraisal of the stimuli presented, as opposed to Nummenmaa et al. (2006), and is arguably more complex than free viewing (which was the task given in Calvo & Lang, 2004; Nummenmaa et al., 2006; and Becker & Detweiler-Bedell, 2009) or than asking a participant to saccade to a given location (Nummenmaa et al., 2009). Additionally, in our experiment, 15 s were given to the participants to explore and memorize each scene. This is a longer display time than what has been typically used in previous eye movement research (3 s in Calvo & Lang, 2004; Nummenmaa et al., 2006; 8 s in Alpers, 2008; and 4 s in Becker & Detweiler-Bedell, 2009). It is possible that with a shorter display time, allocation of attention has to be rushed in order to extract the most relevant information from the scenes. This may increase the role of early attentional processes. The measures of attentional capture by emotional stimuli might, therefore, be less sensitive in our design for this reason.

Finally, in previous paradigms, all stimuli presented in an experiment or block had many structural and semantic similarities, facilitating anticipation and expectation from participants. In the search paradigms cited earlier, 2 × 2 or 3 × 3 matrices of images were used. No more than four semantic categories of objects were used in a single block, with a direct link between semantic category and affect. For example, in Öhman et al. (2001), all inanimate objects (i.e., mushrooms and flowers) were fear-irrelevant, whereas all animals (i.e., spiders and snakes) were fear-relevant. In Calvo and Lang (2004), Nummenmaa et al. (2006, 2009), and Calvo et al. (2008), on every trial, one image was presented in each hemifield. In Calvo and Lang (2004) and Nummenmaa et al. (2006), all images representing inanimate objects were neutral controls, whereas up to two thirds of the images representing people were emotional, enabling participants to expect an emotional content in images representing people. In Becker and Detweiler-Bedell (2009), four faces were presented peripherally. The emotional expressions were limited to neutral, fearful, and happy. In our paradigm, scenes had different layouts and contents, whereas target locations and nature varied for each stimulus. We think that these differences considerably reduced expectation and anticipation effects from participants.

Considering the points discussed above, our results make sense when put in the context of competition for limited resources. It has been observed with fMRI that enhanced activation of the amygdala and visual areas by emotional faces (vs. neutral faces) was only present when the faces were attended (Pessoa, McKenna, Gutierrez, & Ungerleider, 2002). Pessoa and Ungerleider (2004) interpreted these results in terms of limited attentional resources: Task-irrelevant emotional faces are processed only if sufficient resources are available. The researchers went further, hypothesizing that attention is a *sine qua non* for processing emotional faces. Event-related potential (ERP) recordings have led to similar observations, using facial stimuli (Holmes, Vuilleumier, & Eimer, 2003) and IAPS pictures (Schupp et al., 2007).

Our results, seemingly contradicting previous research on attention and emotional stimuli, are in line with scene-perception data, in which the attraction of attention by semantically discrepant objects may depend on the availability of attentional processing resources, which in turn may directly depend on stimulus complexity. In our case, the task was demanding and stimuli were highly complex and cluttered, reducing expectation and anticipation effects from the participants. In light of the capacity-limited attentional resources view, these conditions may be sufficient to prevent an earlier attentional shift toward emotional items in realistic scenes.

Conclusion

We conducted a study assessing the capture of overt attention by emotional stimuli embedded within a complex scene. In contrast with previous research on eye movement using emotional stimuli and sparser displays, we found that emotional targets did not attract attention more than neutral targets in natural scenes. However, once fixated, emotional targets held attention for a longer time. By making participants rate the targets for valence and arousal, we eliminated the hypothesis that our targets had a null emotional impact. We also controlled for low-level pop-out artifacts by comparing targets' visual saliencies outputted by a computational model (Itti & Koch, 2000). We explained the absence of an emotional pop-out effect by arguing that because of stimulus complexity, the task is attentionally demanding, preventing parafoveal emotional information from being processed given the limited attentional resources available. Further research on eye movement and emotion should focus on the manipulation of target nature and size, stimulus complexity, task difficulty, initial foveal load, participants' anticipation by manipulating stimulus variability, and also investigate the effects of individual differences, such as trait or state of anxiety.

References

Alpers, G. (2008). Eye-catching: Right hemisphere attentional bias for emotional pictures. *Laterality: Asymmetries of Body, Brain & Cognition, 13*(2), 158–178.

Bar-Haim, Y., Lamy, D., Pergamin, L., Backermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*(1), 1–24.

Becker, M. W., & Detweiler-Bedell, B. (2009). Early detection and avoidance of threatening faces during passive viewing. *The Quarterly Journal of Experimental Psychology, 62*(7), 1257–1264.

Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance, 33*(1), 20–30.

Blanchette, I. (2006). Snakes, spiders, guns, and syringes: How specific are evolutionary constraints on the detection of threatening stimuli? *The Quarterly Journal of Experimental Psychology, 59*(8), 1484–1504.

Calvo, M. G., & Avero, P. (2008). Affective priming of emotional pictures in parafoveal vision: Left visual field advantage. *Cognitive, Affective, & Behavioral Neuroscience, 8*(1), 41–53.

Calvo, M. G., & Lang, P. J. (2004). Gaze patterns when looking at emotional pictures: Motivationally biased attention. *Motivation and Emotion, 28*(3), 221–243.

Calvo, M. G., & Nummenmaa, L. (2007). Processing of unattended emotional visual scenes. *Journal of Experimental Psychology: General, 136*(3), 347–369.

Calvo, M. G., Nummenmaa, L., & Hyönä, J. (2008). Emotional scenes in peripheral vision: Selective orienting and gist processing, but not content identification. *Emotion, 8*(1), 68–80.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3), 15.

Cave, K., & Batty, M. (2006). From searching for features to searching for threat: Drawing the boundary between preattentive and attentive vision. *Visual Cognition, 14*(4), 629–646.

Coelho, C., Cloete, S., & Wallis, G. (2010). The face-in-the-crowd effect: When angry faces are just cross(es). *Journal of Vision, 10*(7), 14.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52*(4), 317–329.

Flykt, A. (2005). Visual search with biological threat stimuli: Accuracy, reaction times, and heart rate changes. *Emotion, 5*(3), 349–353.

Flykt, A. (2006). Preparedness for action: Responding to the snake in the grass. *The American Journal of Psychology, 119*(1), 29–43.

Fox, E., Griggs, L., & Mouchlianitis, E. (2007). The detection of fear-relevant stimuli: Are guns noticed as quickly as snakes? *Emotion, 7*(4), 691–696.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General, 130*(4), 681–700.

Fox, E., Russo, R., & Dutton, K. (2002). Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cognition & Emotion, 16*(3), 355–379.

Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In R. van Gompel, M. Fischer, W.

Murray, & R. Hill (Ed.), *Eye movements: A window on mind and brain*. Oxford: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243–271.

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition, 17*(6), 1055–1082.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210–228.

Holmes, A., Vuillemier, P., & Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: Evidence from event-related brain potentials. *Cognitive Brain Research, 16*(2), 174–184.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506.

Keil, A., Moratti, S., Sabatinelli, D., Bradley, M. M., & Lang, P. J. (2005). Additive effects of emotional content and spatial selective attention on electrocortical facilitation. *Cerebral Cortex, 15*(8), 1187–1197.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Tech. Rep. No. A-8, University of Florida, Gainesville, FL.

Le Doux, J. E. (1995). In search of an emotional system in the brain: Leaping from fear to emotion and consciousness. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1049–1061). Cambridge, MA: MIT Press.

Lipp, O. V., Derakshan, N., Waters, A. M., & Logies, S. (2004). Snakes and cats in the flower bed: Fast detection is not specific to pictures of fear-relevant animals. *Emotion, 4*(3), 233–250.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*(4), 565–572.

Miller, M., Pasik, P., & Pasik, T. (1980). Extrageniculostriate vision in the monkey. VII. Contrast sensitivity functions. *Journal of Neurophysiology, 43*(6), 1510–1526.

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion, 6*(2), 257–268.

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2009). Emotional scene content drives the saccade generation system reflexively. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 305–323.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117*(2), 382–405.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 446–478.

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences, of the USA, 99*(17), 11458–11463.

Pessoa, L., & Ungerleider, L. G. (2004). Neuroimaging studies of attention and the processing of emotion-laden stimuli. *Progress in Brain Research, 144,* 171–182.

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye movements when looking at unusual/weird scenes: Are there cultural differences? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 254–259.

Rodman, H. R., Gross, C. G., & Albright, T. D. (1989). Afferent basis of visual response properties in area mt of the macaque. I. Effects of striate cortex removal. *Journal of Neuroscience, 9*(6), 2033–2050.

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,* 16.

Soares, S., Esteves, F., & Flykt, A. (2009). Fear, but not fear-relevance, modulates reaction times in visual search with animal distractors. *Journal of Anxiety Disorders, 23*(1), 136–144.

Spielberger, C. D., Gorsuch, R., & Lushene, R. (1970). *The State-Trait Inventory (STAI) test manual*. Palo Alto, CA: Consulting Psychology Press.

Storbeck, J., Robinson, M. D., & McCourt, M. E. (2006). Semantic processing precedes affect retrieval: The neurological case for cognitive primacy in visual processing. *Review of General Psychology, 10*(1), 41–55.

Tipples, J., Young, A. W., Quinlan, P., Broks, P., & Ellis, A. W. (2002). Searching for threat. *The Quarterly Journal of Experimental Psychology: Section A, 55*(3), 1007–1026.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. A. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), *242*–249.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency inuence eye movements when inspecting pictures. *The Quarterly Journal of Experimental Psychology, 59*(11), 1931–1949.

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Consciousness and Cognition, 17*(1), 159–170.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision, 9*(3), 15.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9*(12), 585–594.

Appendix

Stimulus Validation Study

Sixteen participants (10 female) who did not take part in the main experiment were asked to rate the emotional valence and arousal of each target. Participants were asked to rate the target present in each of the images shown. Its location was indicated by a superimposed red dotted circle around the target. Block content was similar as in the main experiment, and stimulus presentation order was randomized for each participant. An instruction sheet explaining the meaning of emotional valence and arousal was given. The head of the participant was fixed in order to ensure similar conditions, such as viewing distance as in the main experiment. Before stimulus onset, a fixation point was displayed for a random duration between 2 and 5 s. The image was then displayed full screen for 10 s, before a dialog box appeared in front of the image, enabling the participant to give their rating. The image was still displayed during the rating, and the participant had no time constraint. Once the rating was over, participants were given their compensation (£3) and signed a receipt.

STAI score differences between participants from the main study (State: $M = 35.58$, $SE = 2.12$; Trait: $M = 38.63$, $SE = 2.27$) and from the validation study (State: $M = 32.88$, $SE = 2.01$; Trait: $M = 36.81$, $SE = 2.35$) were nonsignificant: $t_{state}(30) = .96$ ($p > .3$) and $t_{trait}(30) = .55$ ($p > .5$). On average, mean valence ratings per participant were higher for positive ($M = 2.58$, $SE = .12$) than neutral ($M = 1.09$, $SE = .11$) targets: $t_{P\text{-}Nu}$ $(29.77) = 9.82$ ($p < 10^{-10}$). Mean valence ratings per participant were lower for negative ($M = -2.06$, $SE = .28$) than for neutral targets: $t_{Nu\text{-}Ng}$ $(19.63) = 10.63$ ($p < 10^{-8}$). Mean arousal ratings per participant were higher for both positive ($M = 3.31$, $SE = .25$) and negative ($M = 4.54$, $SE = .39$) than neutral ($M = 2.27$, $SE = .15$): $t_{P\text{-}Nu}$ $(25.14) = 3.59$ ($p < .01$) and $t_{Nu\text{-}Ng}$ $(19.66) = 5.48$ ($p < 10^{-4}$).

Received June 18, 2010
Revision received November 2, 2010
Accepted November 15, 2010 ■