Emotion
2008, Vol. 8, No. 1, 68–80

Copyright 2008 by the American Psychological Association
1528-3542/08/$12.00 DOI: 10.1037/1528-3542.8.1.68

Emotional Scenes in Peripheral Vision: Selective Orienting and Gist Processing, But Not Content Identification

Manuel G. Calvo
University of La Laguna, Spain

Lauri Nummenmaa
MRC Cognition and Brain Sciences Unit, Cambridge, UK, and University of Turku, Finland

Jukka Hyönä
University of Turku, Finland

Emotional-neutral pairs of visual scenes were presented peripherally (with their inner edges 5.2° away from fixation) as primes for 150 to 900 ms, followed by a centrally presented recognition probe scene, which was either identical in specific content to one of the primes or related in general content and affective valence. Results indicated that (a) if no foveal fixations on the primes were allowed, the false alarm rate for emotional probes was increased; (b) hit rate and sensitivity (*A'*) were higher for emotional than for neutral probes only when a fixation was possible on only one prime; and (c) emotional scenes were more likely to attract the first fixation than neutral scenes. It is concluded that the specific content of emotional or neutral scenes is not processed in peripheral vision. Nevertheless, a coarse impression of emotional scenes may be extracted, which then leads to selective attentional orienting or—in the absence of overt attention—causes false alarms for related probes.

Keywords: peripheral and parafoveal vision, emotional, pictorial stimuli, semantic processing

The functional field of view is constrained by the histological properties of the retina, that is, the different density and proportion of rod and cone cells (see Wandell, 1995). In foveate animals, there is a steep drop in visual acuity from the foveal (the central 2° of visual angle) to the parafoveal and the peripheral retina. As stimuli appear farther from the center, the object details fade out, which leads to progressively poorer identification of stimulus content. To overcome these biological constraints and be able to construct an accurate representation of the multiple objects in the visual environment, spatial attention relies on sequential scanning by means of eye movements that bring objects to the fovea. However, given the adaptive importance of emotional stimuli, it is possible that a capacity has evolved to amplify affective visual signals from the peripheral retina, without eye fixations. In other words, is the span of effective vision, or functional field of view, broadened for emotional stimuli? To address this issue, we investigated whether the content of emotional visual scenes is more likely to be identified than that of nonemotional scenes when they are presented in the visual periphery, and the extent to which such an advantage may depend on overt attention to these stimuli.

Prior research has found that, compared with neutral pictures, emotional pictures have facilitated access to the cognitive system when they appear in the center of the visual field (i.e., foveally). Studies using behavioral and physiological measures have revealed enhanced processing of emotional stimuli when they are presented very briefly or even subliminally, and followed by a mask (with an affective priming paradigm, Hermans, Spruyt, De Houwer, & Eelen, 2003; with recognition sensitivity measures, Calvo & Esteves, 2005; with electromyographical assessment, Dimberg, Thunberg, & Elmehed, 2000; with electrodermal assessment, Öhman & Soares, 1998). In addition, studies measuring event-related potentials (ERP) have shown that centrally presented emotional facial expressions elicit early posterior or frontocentral responses usually within 100 to 200 ms poststimulus onset (e.g., Pourtois, Thut, Grave de Peralta, Michel, & Vuilleumier, 2005; see Eimer & Holmes, 2007; Palermo & Rhodes, 2007, for reviews). ERP modulations originating from visual cortices occur later for complex emotional scenes (such as scenes depicting acts of violence) than for faces, starting between 150 ms (Junghöfer, Bradley, Elbert, & Lang, 2001) and 300 ms (Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000) after picture onset. Likewise, neuroimaging studies using positron emission tomography (PET; Lane, Chua, & Dolan, 1999) and functional MRI (fMRI; Sabatinelli, Bradley, Fitzsimmons, & Lang, 2005) have demonstrated enhanced responses in the amygdala and the visual cortices to emotional faces and scenes, relative to neutral stimuli.

This privileged access of emotional stimuli to the visual processing systems in the brain is, presumably, due to their adaptive importance as signals of appetitive or aversive events. These

---

Manuel G. Calvo, Department of Cognitive Psychology, University of La Laguna, Spain; Lauri Nummenmaa, MRC Cognition and Brain Sciences Unit, Cambridge, UK, and University of Turku, Finland; and Jukka Hyönä, Department of Psychology, University of Turku, Finland.

This research was supported by Grant SEJ2004–0049, from the Spanish Ministry of Education and Science to Manuel G. Calvo and Grant 119088 from the Academy of Finland to Lauri Nummenmaa. We thank Aapo Nummenmaa for his invaluable help with assessing the low-level physical properties of the images.

Correspondence concerning this article should be addressed to Manuel G. Calvo, Department of Cognitive Psychology, University of La Laguna, 38205 Tenerife, Spain. E-mail: mgcalvo@ull.es

findings are consistent with an evolutionary perspective: For adaptive purposes, the perceptual system is biased toward efficiently discovering threat and harm, as well as benefit and pleasure (see Lang, Bradley, & Cuthbert, 1990). Automatic (i.e., quick and efficient, and even independent of awareness and volition) processing of such stimuli is assumed to be functional in readily recruiting resources to either cope with potentially harmful events or to take advantage of beneficial situations. This would enable the individual to prompt a rapid preservative or protective response, and thus maximize benefit and minimize harm. Most prior research using emotional faces or scenes has focused on the *temporal* component of such an emotional processing enhancement mechanism, that is, the perceptual threshold or the time course of encoding of emotional stimuli. As a complementary approach, studies are needed that examine the *spatial* component, that is, a possible broadening of the functional field of view when detecting emotional scenes. The role of spatial attention has been assessed in many studies on emotional face processing using ERP measures (see Eimer & Holmes, 2007; Palermo & Rhodes, 2007; for reviews). In contrast, this issue has been addressed to a much lesser extent in studies on emotional scene processing (with ERP measures, Keil, Moratti, Sabatinelli, Bradley, & Lang, 2005; with behavioral measures, Calvo & Nummenmaa, 2007). Also, in most of the previous studies neutral and emotional stimuli were presented one at a time. A complementary way to examine the preferential processing of emotional stimuli, that is, their *relative* accessibility to encoding, in comparison with neutral stimuli, involves concurrent presentation of these two types of stimuli. In the current study, we extended prior research by addressing both the spatial attention and the relative accessibility issues.

The specific research question is, therefore, whether the content of emotional visual scenes is analyzed when these appear in peripheral areas of the visual field—outside of the focus of overt attention (without eye fixations)—and whether this involves preferential processing of emotional scenes over concurrent neutral scenes. There are some arguments for predicting such a broadening of the visual field. First, given that most of the objects in a rich visual environment normally fall beyond the limits of our foveal vision, adjustments should take place in the breadth of the visual field to maximize the probability that events with special adaptive significance are detected. Second, results obtained with the dot-probe procedure (Mogg, Bradley, Miles, & Dixon, 2004), the visual search task (Blanchette, 2006), and the exogenous cuing paradigm (Fox, Russo, Bowles, & Dutton, 2001), suggest that emotional pictures (including faces, objects, and scenes) may receive preferential attention outside of foveal vision. These studies show facilitated processing of emotional images compared with neutral images that appear concurrently but away from a central fixation point. Nevertheless, these studies were not designed to investigate the role of stimulus spatial eccentricity on content identification, as the tasks normally allowed fixations to be made on the stimuli following their parafoveal or peripheral presentation.

Third, findings of selective orienting of eye movements toward emotional scenes presented concurrently with neutral scenes in the visual periphery are particularly relevant to the issue under investigation. In the Calvo and Lang (2004) and the Nummenmaa, Hyönä, and Calvo (2006) studies, pairs of one emotional and one neutral scene (or two neutral scenes), were presented simultaneously for 3 s, with their inner edges between 4.0° and 9.2° away from a central fixation point. Participants were instructed to respond whether the two scenes were of the same valence or not. Calvo and Lang found that the probability of first fixation and the viewing time during the first 500 ms were greater for both pleasant and unpleasant scenes than for the paired neutral scenes. This early selective orienting to emotional stimuli was replicated by Nummenmaa et al. Furthermore, in the Nummenmaa et al. study, the emotional picture of each pair was more likely to be fixated first and gazed longer than the neutral picture during the first-pass viewing (i.e., the time elapsed since the eyes landed first on a picture until they exited it) not only during free viewing conditions, but also when the participants were instructed to look at the neutral picture first and to keep fixating it. In contrast, after the first 650 to 700 ms, the participants were able to control their gaze and comply with the instructions (i.e., look at the neutral picture, when requested) for the rest of the 3-s display time. This was interpreted as an early involuntary attention capture by emotional content, which requires effort and time to be counteracted voluntarily.

The selective orienting to peripheral emotional scenes suggests that something of them is perceived by covert attention that brings about an overt eye fixation to them. The question is whether this "something" involves (a) meaningful and specific content, (b) global or "gist" information (Underwood, 2005), or (c) whether the attraction of saccades is determined merely by low-level visual features. According to theories based on saliency mapping models (Parkhurst, Law, & Niebur, 2002), visual attention is guided to the scene or scene region that has the highest perceptual weight within the saliency map, and perceptual weight of each location is initially determined by factors such as luminance, color, texture, and so forth. Only after a scene object (or its nearby regions) has been fixated, semantic factors begin to influence the saliency map. However, the results of the Calvo and Lang (2004) and Nummenmaa et al. (2006) studies cannot be accounted for by the saliency mapping theories, as the emotional and neutral scenes were comparable in luminance, contrast, and color saturation. An alternative hypothesis that orienting is driven by peripheral processing of emotional content was not examined by Calvo and Lang or Nummenmaa et al., as these authors did not take any measure of scene mental representation. The possibility that semantic content of (nonemotional) scenes is processed outside of foveal vision has recently received support (De Graef, 2005; Underwood, 2005). The gist (or scene category) can be extracted peripherally, without the identification of individual objects in the scene (Gordon, 2004; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001). Accordingly, an interesting issue to be investigated is whether emotional valence is sufficiently salient to be detected peripherally, and what type of meaning (detailed vs. gist) can be extracted prior to overtly attending to the scenes.

To address these issues, we examined in the current study (a) whether the specific content of emotional scenes is more likely to be identified than that of neutral scenes when they are presented in the visual periphery, and (b) whether this advantage involves genuine peripheral processing or, rather, depends on overt attentional orienting. To this end, we presented pairs of prime pictures (one neutral, one emotional) with a visual angle of 5.2° away from a central fixation point for a short duration (150 ms) to prevent any fixations on the target pictures (Experiment 1). Given that the minimum saccade latency (i.e., the time to initiate an eye movement from the current fixation to another stimulus) is about 150 ms (see Rayner, 1998), this short presentation time was deemed to be adequate to assess peripheral processing. In subsequent experiments, pairs of emotional and neutral pictures were presented peripherally either for 300 ms (Experiment 2A) or for 450 ms (Experiment 2B), to allow time for an eye fixation on only one picture of the pair; or for 900 ms (Experiment 2C), to allow a fixation on both pictures. If genuine peripheral processing of emotional content occurs, an advantage in the identification of emotional over neutral scene content will appear in the 150-ms condition. In contrast, if foveal attention is required for content identification, such an advantage will appear only when scenes can be overtly attended, that is, in the longer display conditions. The aforementioned distinction between truly peripheral processing (i.e., in the absence of eye fixations) and foveal processing (i.e., with eye fixations) was directly tested by means of eye-movement monitoring (Experiment 3).

The presentation of the prime picture pairs was immediately followed by a recognition test, in which we used two types of probe pictures to assess the type of information acquired from the primes. On half of the trials, the probe had the same specific content and affective valence as one of the primes (the *identical* condition). On the other half, the probe was different in specific content from the primes, but it was of the same affective valence and general topic as one of the primes (the *related* condition). In all cases, the prime and the probe differed in color (full color vs. grayscale), size (smaller vs. larger), and spatial orientation (probes were mirror images of the primes), to minimize the contribution of perceptual features and maximize the role of meaning in recognition. Evidence for the processing of specific semantic content would show up as more accurate (i.e., higher hit rate and detection sensitivity) and faster recognition of identical probes for emotional than for neutral scenes. In contrast, if the advantage in the processing of emotional primes is only due to extracting a general impression about its affective significance, there will be more false alarms (rather than hits) for related emotional probes than for related neutral probes, accompanied by a poor discrimination between the identical and the related emotional scenes (as both shared the same emotional valence).

Experiment 1

Pairs of emotional-neutral prime pictures were presented peripherally for 150 ms to either side of a central fixation point, followed by a mask and a probe picture for recognition. The probe was either identical in content to one of the prime pictures (target-present trials) or related in gist and affective valence (target-absent, or catch trials). The related probes in the catch trials were used to assess false alarms and estimate detection sensitivity.

Method

Participants. The 28 participants (20 women) in Experiment 1 and those in the following experiments were undergraduate psychology students (between 18 and 25 years of age) at La Laguna University, who participated for course credit. They were informed that they would be presented with photographs, some of which could be pleasant or unpleasant in content.

Stimuli. In all the experiments, we used the same 128 stimuli. For the target-present trials, 64 digitized color photographs were presented as target stimuli, of which 32 were of neutral in affective valence (i.e., nonemotional), 16 were unpleasant and 16 were pleasant. Each emotional picture was randomly paired with a neutral picture on each trial. All target pictures portrayed people, either (a) in a variety of daily activities (neutral scenes), or (b) suffering harm (unpleasant scenes: violent attacks, seriously injured or dead people, or expressions of pain, crying, or despair), or (c) enjoying themselves (pleasant scenes: sports and recreational activities, heterosexual erotic couples, expressions of love, happy families. and babies). For the catch trials, an additional group of 64 pictures were selected. Each of these pictures was matched with one of the target pictures in topic, presence of people, valence, and arousal. Thus, the target and the matched scenes were similar in general content and emotionality, but their specific content was different. All the target stimuli and most of the matched stimuli were selected from the International Affective Picture System (IAPS; Center for the Study of Emotion & Attention, 2005)[1]. Figure 1 shows examples of the target and matched scenes.

When presented as primes, all pictures were in their original colors and spatial orientation, but reduced in size. Each prime picture subtended a visual angle of 13.3° (width: 11.2 cm) by 10.9° (height: 9.2 cm). A central fixation point was located horizontally between the two pictures, at 5.2° of visual angle (4.33 cm) from the inner edge of each picture. Accordingly, the distance from the fixation point to the center of each picture was 11.8°. When presented for recognition, as probes, all pictures were in grayscale, increased in size (35.8° by 26.9°), and their left-right orientation was mirror reversed.

Apparatus and procedure. The pictures were displayed on a SVGA 17” monitor with a 100-Hz refresh rate, connected to a Pentium III computer. The E-Prime software (Schneider, Eschman, & Zuccolotto, 2002) controlled stimulus presentation and response collection. Participants had their head positioned on a chin and forehead rest, with their eyes located at a distance of 48 cm from the center of the screen. Response accuracy and latency were collected through presses on specified keys of the computer keyboard.

Participants were informed that they would be presented with pairs of photographs, and that they should fixate an initial central cross until two color photographs appeared, one of which could be presented for immediate recognition. It was made clear that the

---

[1] The IAPS numbers for the target scenes and the corresponding matched scenes (in parentheses) were: (a) neutral pictures: 2037 (2357), 2190 (2493), 2191 (2191.1), 2220 (2200), 2221 (5410), 2270 (9070), 2272 (2272.1), 2312 (2312.1), 2383 (2372), 2393 (2393.1), 2394 (2305), 2396 (2579), 2397 (2397.1), 2512 (2491), 2513 (2513.1), 2560 (2560.1), 2575 (2575.1), 2593 (2593.1), 2594 (2594.1), 2595 (2595.1), 2598 (2598.1), 2635 (2635.1), 2745.1 (2745.2), 2749 (2749.1), 2840 (2410), 2850 (2515), 2870 (2389), 7493 (2102), 7496 (7496.1), 7550 (7550.1), 7620 (7620.1), and 9210 (9210.1); (b) unpleasant pictures: 2399 (2399.1), 2691 (2683), 2703 (2799), 2718 (2716), 2800 (2900), 2811 (6250), 3180 (3181), 3225 (3051), 3350 (3300), 6010 (2722), 6313 (6315), 6550 (6560), 8485 (8480), 9254 (9435), 9410 (9250), and 9423 (9415); and (c) pleasant pictures: 2070 (2040), 2165 (2160), 2540 (2311), 2550 (2352), 4599 (4610), 4647 (4694), 4658 (4680), 4669 (4676), 4687 (4660), 4700 (4624), 5621 (8161), 5831 (5836), 7325 (2332), 8186 (8021), 8200 (8080), and 8490 (8499).

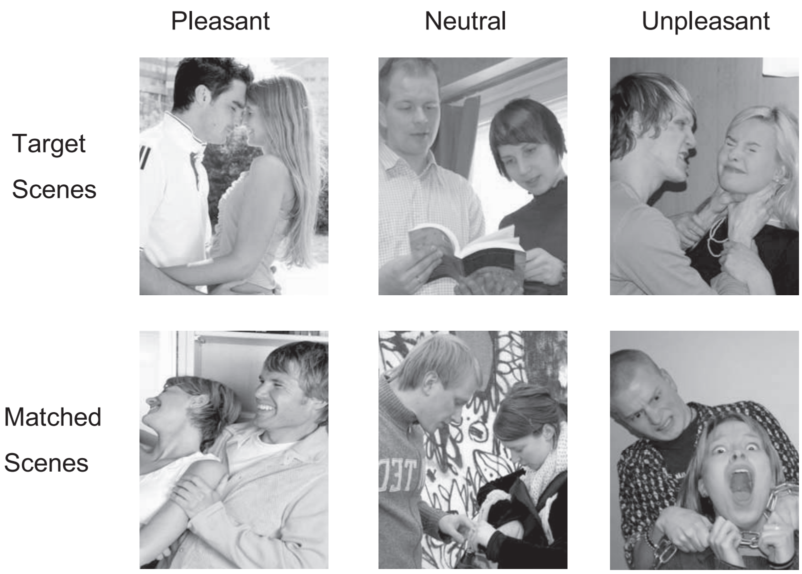

Figure 1. Examples of target and matched (related in content and affect) neutral and emotional scenes. Examples of original pictures (not taken from the IAPS for copyright reasons). The original pictures were presented in landscape format.

photograph for recognition would increase in size, change from color to grayscale, and change in left-right orientation. The participant was to respond, as soon as possible, whether this modified photograph had, nevertheless, the same content ("the same people doing the same things") as one of the two previously paired photographs, by pressing a "Yes" key (L) or a "No" key (D). Twelve practice trials were followed by 128 experimental trials.

Figure 2 shows the sequence of events on each trial. A trial started with a fixation cross for 750 ms, followed by a 100-ms blank interval, after which the cross reappeared for 150 ms. This "flashing" of the cross served to capture the viewer's attention on the central location just before the prime pictures appeared. Next, a pair of prime pictures (one emotional, one neutral) was displayed peripherally for 150 ms. Following prime offset, a mask (a random combination of colors) was presented for 500 ms. Finally, a probe picture (either a target or a matched item) appeared for recognition until the participant responded.

Each participant saw each of the 128 experimental pictures only once as a probe in the recognition test. The target probe appeared in one block (either first or second) and the corresponding matched probe appeared in the other block (second or first). Participants saw each picture as a prime four times (twice in each block), each time paired with a different picture. On two occasions, the to-be-probed picture appeared in the left visual field; on another two, in the right field. In both visual fields, the prime preceded an identical probe (i.e., target trials) and a related probe (i.e., catch trials) once. On target trials, the probe was identical in specific content, although different in color, size, and orientation to one of the primes. On catch trials, the probe was related in topic and emotionality to one of the primes, although different in form (color, size, and orientation) and specific content. This produced a within-subjects factorial combination of 3 (emotional valence of the prime: unpleasant vs. neutral vs. pleasant) × 2 (prime-probe relationship: identical vs. related in content) × 2 (visual field of the prime: left vs. right). The trials were presented in random order.

Results

Characteristics of the stimuli. We compared the three groups of pictures (unpleasant, neutral, and pleasant) with respect to their rated affective valence and arousal, as well as low-level image properties. Table 1 shows the mean scores for these variables.

Valence and arousal ratings for each picture (in a 9-point scale) have been obtained in norming studies (Lang, Bradley, & Cuthbert, 2005). Valence (unpleasantness vs. pleasantness) reflects the dominant motive system activated (avoidance or approach). Arousal reflects the intensity of motive system activation, from calm to tension. A Valence Category (unpleasant vs. neutral vs. pleasant picture) × Relatedness (target vs. matched pictures) analysis of variance (ANOVA) on valence rating scores yielded a strong effect of category, $F(2, 122) = 811.15$, $p < .0001$, with significant differences between all three categories (all post hoc ps $< .0001$, after Bonferroni-corrected multiple comparisons). The effects of relatedness, $F(1, 122) = 0$ and the interaction, $F(2, 122) = 1.34$, $p = .27$ were not significant. For arousal ratings, the effect of valence category, $F(2, 122) = 32.28$, $p < .0001$, revealed higher arousal scores for the unpleasant and the pleasant stimuli than for the neutral stimuli (both ps $< .0001$), but no difference between the pleasant and unpleasant stimuli. The effects of relatedness and the interaction were not significant, both $Fs < 1$. The lack of any interaction is important because it rules out the hypothesis that improved recognition of emotional scenes could be due to a closer match of valence and arousal between emotional primes and probes, as compared with neutral primes and probes.

Using Adobe Photoshop 6.0, we computed color saturation (red, green, and blue) values for each target picture, and with Matlab (The Mathworks, Natick, MA) we calculated basic image statistics such as mean luminance, variance of luminance, root mean square (RMS) contrast, kurtosis, skewness, and energy (see Kirchner & Thorpe, 2006). The luminance and contrast of some of the original pictures were readjusted to make the three valence categories

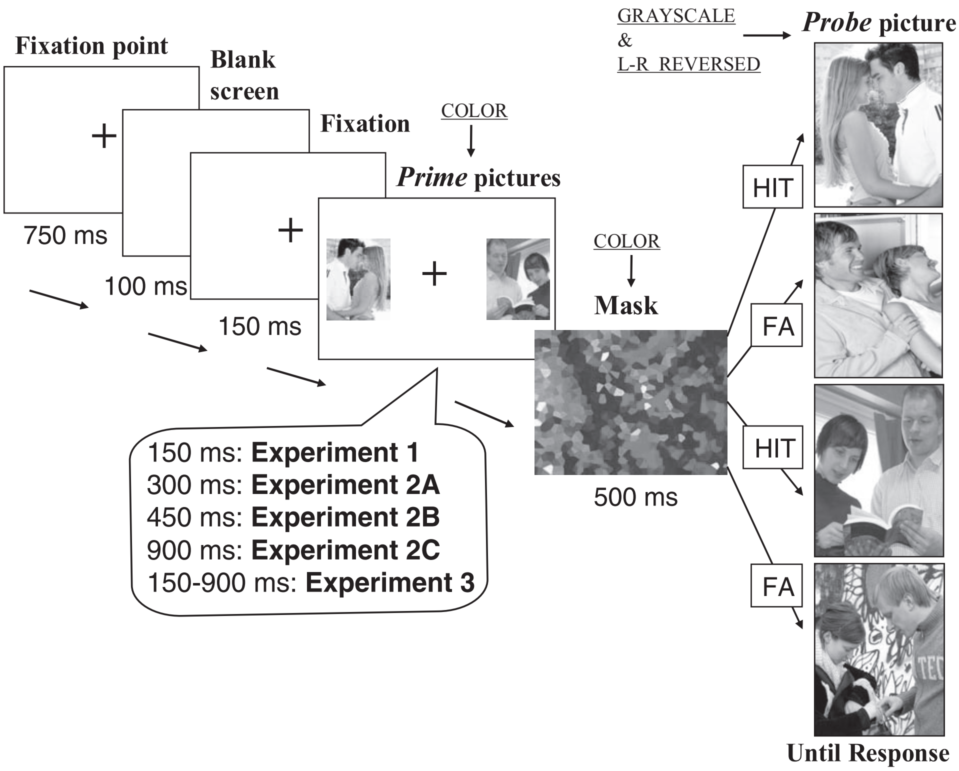

Figure 2. Sequence of events within a trial in Experiments 1–3. Only one probe picture appeared on each trial, either the emotional or neutral target scene (hit), or the emotional or neutral matched scene (false alarm). In Experiment 3, a drift-correct circle appeared as the fixation point (instead of the flashing cross), until the participant fixated on its center.

comparable. The one-way (valence) ANOVAs showed no significant differences in any of these image characteristics, all Fs $\leq$ 1.30, $p \geq .28$; for energy: $F(2, 61) = 3.13$, $p \geq .082$, *ns*, after corrections for multiple contrasts. These analyses rule out the possibility that effects of valence on attentional orienting would be merely due to low-level visual factors. Furthermore, potential differences between the target and the matched pictures for each valence category were analyzed by means of two-way (Valence ×

Table 1
Mean Scores of the Stimulus Characteristics of the Emotional and the Neutral Stimuli Used in Experiments 1 to 3

	Type of scene		
Characteristic	Unpleasant	Neutral	Pleasant
Valence rating (target items)	2.26 (0.52)^{a}	5.19 (0.59)^{b}	7.68 (0.37)^{c}
Arousal rating (target items)	5.96 (1.36)^{a}	4.23 (0.56)^{b}	5.80 (1.57)^{a}
Valence rating (matched items)	2.47 (0.65)^{a}	5.02 (0.51)^{b}	7.65 (0.41)^{c}
Arousal rating (matched items)	5.79 (1.40)^{a}	4.30 (0.49)^{b}	5.54 (0.43)^{a}
Red saturation	111 (21)	114 (22)	125 (38)
Green saturation	90 (18)	91 (14)	92 (24)
Blue saturation	79 (16)	80 (24)	78 (32)
Luminance (average)	100 (14)	104 (16)	107 (20)
Luminance (variance)	70 (11)	70 (12)	27 (13)
RMS	0.70 (0.14)	0.67 (0.11)	0.69 (0.18)
Skewness	0.64 (0.39)	0.55 (0.53)	0.35 (0.63)
Kurtosis	2.44 (0.50)	2.52 (1.17)	2.36 (1.08)
Energy ($\times 10^{-5}$)	7.285 (3.189)	8.721 (2.722)	6.876 (1.829)

Note. Standard deviations are in parentheses. A different subscript indicates significant differences between scene categories (if two scores share an identical subscript, they are equivalent). RMS = root mean square.

Relatedness) ANOVAs on each of these characteristics. The effects of relatedness were not statistically significant, all Fs < 1, except for RMS: $F(1, 122) = 2.27$, $p = .14$, *ns*. Most important, the interaction was not significant, all Fs < 1, except for luminance variance: $F(2, 122) = 2.79$, $p \geq .073$, *ns*, after corrections for multiple contrasts. This rules out the hypothesis that improved discrimination between presented (target) and nonpresented (matched) pictures for one valence category, relative to the others, could be due merely to more low-level visual differences between the primes and the probes for that valence category.

Recognition data. The probability of hits (PH; i.e., correct "yes" responses to probes in target-present trials) and false alarms (PFA; i.e., incorrect "yes" responses to probes in target-absent trials) in recognition performance were converted to an index of sensitivity or discrimination, A' (see Snodgrass & Corwin, 1988); where $A' = 0.5 + (\mathrm{PH} - \mathrm{PFA}) * (1 + \mathrm{PH} - \mathrm{PFA})/(4 * \mathrm{PH}) * (1 - \mathrm{PFA})$. A' scores vary from low to high sensitivity in a 0 to 1 scale. In addition, a response bias index (B'') was computed (see Snodgrass & Corwin, 1988), according to the formula: $B'' = ((1 - \mathrm{PH}) * (1 - \mathrm{PFA}) - (\mathrm{PH}) * (\mathrm{PFA}))/((1 - \mathrm{PH}) * (1 - \mathrm{PFA}) + (\mathrm{PH}) * (\mathrm{PFA}))$. B'' scores vary from -1.0 to $+ 1.0$: A positive value indicates a strict or conservative criterion, a zero value indicates a neutral criterion, and a negative value indicates a liberal or lenient criterion.

The mean scores for each dependent variable in Experiment 1 are presented in Table 2. A series of 3 (emotional valence: unpleasant vs. neutral vs. pleasant) $\times$ 2 (visual field: left vs. right) repeated measures ANOVAs were conducted for the probability of hits and false alarms, sensitivity, and reaction times for hits. For all the experiments, a significant main effect of valence was examined further by Bonferroni-corrected post hoc comparisons.

The effect of emotional valence was not significant for the probability of hits, $F(2, 54) = 1.67$, $MSE = 0.027$, sensitivity, or response bias scores, $Fs < 1$. Furthermore, the hit rates did not differ from the chance level (i.e., .5; all one-sample t tests <1.2, $p > .20$). However, the effect of valence was reliable for the false alarm rate, $F(2, 54) = 9.33$, $MSE = 0.014$, $p < .001$, with more false alarms for both unpleasant and pleasant scenes (both ps $< .01$) than for neutral scenes. Reaction times were not reliably affected by any factor.

Discussion

When the primes were presented peripherally for 150 ms, the probe recognition hit rate was at the chance level for all types of scenes. Moreover, there was no advantage of emotional over neutral stimuli in either hit rates or sensitivity scores. This implies that no information about the specific content of scenes (i.e., the *identity* of the portrayed objects, people, or actions) was acquired in peripheral vision. In contrast, with a parafoveal presentation of the same stimuli, Calvo and Lang (2005) found that emotional scenes were more likely to be recognized than neutral scenes, as revealed by both hits and sensitivity scores. Furthermore, Calvo (2006) compared a parafoveal and a peripheral condition in a within-subjects design, and observed an interaction between emotional valence and eccentricity of the primes: The recognition advantage for the emotional stimuli occurred in the parafoveal but not in the peripheral presentation condition.

Table 2
Mean Probability of Hits and False Alarms (FAs), Sensitivity Scores (A′), Response Bias Scores (B″), and Reaction Times (RT) for Hits As a Function of Emotional Valence of Scenes in Experiment 1

Valence	Hits	FAs	A'	B''	RT LVF	RT RVF
Unpleasant	.558	$.298^{a}$	.684	.351	963	988
Neutral	.506	$.208^{b}$	.716	.425	937	955
Pleasant	.550	$.284^{a}$	.698	.363	967	1,016

Note. Experiment 1 = 150-ms stimulus display, paired prime pictures. A different subscript indicates significant differences between valence categories. Hits referred to are in the left visual field (LVF) versus the right visual field (RVF), respectively.

An additional new finding of Experiment 1 is the significant effect of emotional valence on false alarms, which in contrast, did not appear in the previous studies using parafoveal stimulus presentation (Calvo, 2006; Calvo & Lang, 2005). Under peripheral prime presentation, there were more false alarms for emotional than for neutral stimuli. A possible explanation is that a gist impression was obtained from the emotional scenes in peripheral vision, rather than a detailed representation of object identities. This reasoning is based on the fact that false alarms represent incorrect "yes" responses to probe pictures that had not been presented as primes, but that were related in general content and emotional valence to the presented primes. If participants obtained only a global impression of the primes, this would lead them to confound a probe with a nonpresented, but related prime, and to wrongly accept the probe as presented. If, instead, the specific content of the prime stimuli had been processed, this should have been reflected in an increased hit rate and sensitivity for the target probes, which was not the case. The alternative interpretation that the higher false alarm rate for emotional than for neutral scenes could be due to response biases is not supported, as there were no differences in the response bias index (B'') as a function of scene valence category.

Two predictions can be made based on the aforementioned hypothesis that a coarse impression of peripheral emotional scenes can be extracted, but not a detailed semantic representation of its components. First, only when there is sufficient time to make an eye fixation on the prime stimulus, will specific information be acquired from peripherally presented emotional scenes, and to a greater extent than from neutral scenes. This implies that there is no genuine specific semantic processing of peripherally presented scenes. To obtain an accurate semantic representation, the scenes need to be overtly attended. To examine this prediction, we increased the prime stimulus duration in Experiments 2A, 2B, and 2C. Second, eye movements will be initially directed to the emotional rather than to the neutral picture of each pair. The reason is that, if a coarse gist impression of emotional content can be obtained peripherally, this will attract overt attention first to the emotional scene. This prediction was assessed in Experiment 3 using eye-movement monitoring.

Experiments 2A, 2B, and 2C

Emotional-neutral picture pairs were presented peripherally for 300, 450, or 900 ms (Experiments 2A, 2B, and 2C, respectively), to examine the extent to which overt attention (i.e., foveal fixation) is necessary to identify the specific content of peripheral visual

Table 3

Mean Probability of Hits and False Alarms (FAs), Sensitivity Scores (A′), Response Bias Scores (B″), and Reaction Times (RT) for Hits As a Function of Emotional Valence of Scenes in Experiments 2A, 2B, and 2C

Valence	Hits	FAs	*A′*	*B″*	RT LVF	RT RVF
Experiment 2A						
Unpleasant	.787^{a}	.259^{a}	.832	.136	851	897†
Neutral	.683^{b}	.204^{b}	.826	.223	821	918†
Pleasant	.839^{a}	.274^{a}	.861	.129	841	920†
Experiment 2B						
Unpleasant	.839^{a}	.212	.886^{a}	.078	851	905†
Neutral	.674^{b}	.200	.821^{b}	.158	839	918†
Pleasant	.845^{a}	.236	.881^{a}	.099	860	923†
Experiment 2C						
Unpleasant	.869	.158	.915	.011	883	896
Neutral	.819	.147	.897	.102	867	899
Pleasant	.879	.148	.923	.023	874	902

Note. A different subscript indicates significant differences between valence categories; a dagger (†) indicates significant reaction time differences between the left and the right visual field. Hits referred to are in the left visual field (LVF) versus the right visual field (RVF), respectively. Experiment 2A = 300-ms stimulus display (paired prime pictures); Experiment 2B = 450-ms stimulus display (paired prime pictures); Experiment 2C = 900-ms stimulus display (paired prime pictures).

scenes. Given the estimation of minimum saccade latency of 150 ms and a modal fixation duration for pictorial stimuli of about 225 ms (see Henderson & Hollingworth, 1998), we assumed that: (a) in the 300-ms condition, only one very short, incomplete fixation would be possible on only one picture of the pair; (b) in the 450-ms condition, a single complete fixation could take place only on one picture; (c) in the 900-ms condition, at least one complete fixation could be made on both pictures. Accordingly, if emotional scene content attracts covert attention peripherally but overt attention is required for the identification of emotional scene content, an advantage for emotional relative to neutral scenes will occur when there is enough time to look at only one picture, that is, in the 450-ms condition. In contrast, in the 300-ms condition, the time available for foveal inspection of the scene may be insufficient for the processing of specific semantic content, which will lead to increased false alarms, rather than more hits and higher sensitivity for emotional than for neutral scenes. In the 900-ms condition, there should be no differences in any recognition measure as a function of emotional valence, as both prime pictures could be overtly attended.

Method

There were 24 (17 female), 24 (16 female), and 24 (18 female) participants in Experiments 2A, 2B, and 2C, respectively. The method was identical to that of Experiment 1, except that the prime pictures were presented for 300, 450, or 900 ms, instead of 150 ms.

Results

Separate analyses for each experiment. The mean scores for each dependent variable are shown in Table 3. Initially, 3 (Emotional Valence) × 2 (visual field) repeated-measures ANOVAs were conducted for each experiment separately. In Experiment 2A (300-ms display), the effect of emotional valence was significant for the probability of hits, $F(2, 46) = 12.95$, $MSE = 0.022$, $p < .0001$, and false alarms, $F(2, 46) = 4.04$, $MSE = 0.017$, $p < .025$, but not for sensitivity scores, $F(2, 46) = 2.57$, $MSE = 0.007$, $p = .09$, or response bias scores, $F < 1$. Post hoc comparisons indicated that there were more hits, but also more false alarms, for both pleasant and unpleasant scenes than for neutral scenes ($ps < .05$). Reaction times for hits varied as a function of visual field, $F(1, 23) = 8.47$, $MSE = 23{,}315$, $p < .01$, with faster responses when the probe pictures were presented as primes in the left ($M = 838$ ms) than in the right field ($M = 912$ ms).

In Experiment 2B (450-ms display), the effect of emotional valence was significant for the probability of hits, $F(2, 46) = 19.44$, $MSE = 0.015$, $p < .0001$, and sensitivity scores, $F(2, 46) = 13.14$, $MSE = 0.005$, $p < .0001$, but not for false alarms or response bias scores, $Fs < 1$. The hit rate and sensitivity scores were higher for both pleasant and unpleasant scenes than for neutral scenes ($ps < .01$). Reaction times for hits were shorter when the probe pictures were presented as primes in the left ($M = 850$ ms) than in the right field ($M = 915$ ms), $F(1, 23) = 4.36$, $MSE = 35{,}250$, $p < .05$.

In Experiment 2C (900-ms display), there were no reliable effects on the probability of hits, false alarms, sensitivity, response bias scores, or reaction times (all $ps > .15$).

Pooled analyses of four experiments. To examine the role of overt attention, Experiments 1, 2A, 2B, and 2C were combined in a 4 (prime display duration: 150 vs. 300 vs. 450 vs. 900 ms) × 3 (valence) × 2 (visual field) ANOVA. For the probability of hits, main effects of display duration, $F(3, 96) = 32.20$, $MSE = 0.087$, $p < .0001$, and valence, $F(2, 192) = 29.02$, $MSE = 0.023$, $p < .0001$, were qualified by their interaction, $F(6, 192) = 2.78$, $MSE = 0.023$, $p < .025$. The hit rate was lower in the 150-ms condition ($M = .538$) than in all the other conditions ($M = .770$ vs. .786 vs. .855, for the 300-, 450-, and 900-ms conditions, respectively), and lower in the 300- than in the 900-ms condition (all $ps < .05$). The interaction reflects the fact that the hit rate was greater for the pleasant and the unpleasant scenes than for the neutral scenes in the 300- and the 450-ms condition, but not in the 150- and the 900-ms condition, as indicated by the separate analysis of each experiment (see above).

The false alarm rate was affected by display duration, $F(3, 96) = 4.85$, $MSE = 0.075$, $p < .01$, and valence, $F(2, 192) = 8.58$, $MSE = 0.015$, $p < .0001$, with no interaction ($p = .11$). There were more false alarms in the 150- and 300-ms conditions ($M = .263$ and .246) than in the 900-ms condition ($M = .151$), as well as more false alarms for both unpleasant and pleasant scenes ($M = .231$ and .236) than for neutral scenes ($M = .190$; all $ps < .05$).

The analyses of the sensitivity index yielded main effects of display duration, $F(3, 96) = 39.46$, $MSE = 0.034$, $p < .0001$, and valence, $F(2, 192) = 2.99$, $MSE = 0.010$, $p = .053$, which were qualified by their interaction, $F(6, 192) = 2.16$, $MSE = 0.010$, $p < .05$. Sensitivity scores were lower in the 150-ms condition ($M = .699$) than in all the other display conditions ($M = .840$ vs. .865 vs. .912, for the 300-, 450-, and 900-ms conditions, respectively), and lower in the 300- than in the 900-ms condition (all $ps < .05$). The interaction reflects the fact that sensitivity was higher for the pleasant and the unpleasant scenes than for the neutral scenes in the 450-ms condition, but not in the other exposure conditions, as indicated by the separate analysis each experiment (see above).

Finally, for reaction times of correct responses, there was a marginal effect of display duration, $F(3, 96) = 2.56$, $MSE = 130{,}313$, $p = .060$, and a highly reliable effect of visual field, (1, 96) = 13.52, $MSE = 24{,}865$, $p < .0001$. Latencies tended to be longer in the 150-ms ($M = 971$ ms) than in the 300-, 450-, and 900-ms displays ($M = 873$ vs. 883 vs. 887 ms, respectively), although the contrasts were not statistically significant after Bonferroni correction for multiple comparisons. Latencies were shorter for probes when the prime was presented to the left than to the right visual field ($M = 879$ vs. 927 ms).

Discussion

The results of Experiments 2A, 2B and 2C, and their comparison with Experiment 1, revealed two main findings. The first is concerned with the role of overt attention to the peripherally presented stimuli. Identification of stimulus content, as shown by the sensitivity index, improved in all three display conditions that were assumed to allow foveal fixations on the pictures (i.e., the 300-, 450-, and 900-ms displays), in comparison with the condition that was assumed not to allow overt attention (i.e., the 150-ms display). A similar improvement occurred also for the hit rate, which was at chance level in the 150-ms condition and above chance in all the other exposure conditions. This implies that overt attention is required for the processing of specific semantic content of visual stimuli.

The second major finding is concerned with the advantage of emotional stimuli over neutral stimuli. Increased exposure time, and therefore the possibility for foveally attending to a peripheral stimulus, benefited more the identification of emotional scenes than that of the paired neutral scenes only in the 450-ms display. In the 300-ms condition, the hit rate advantage for emotional scenes was accompanied by a higher false alarm rate, which led to no difference in the discrimination index of emotional versus neutral stimuli. In the 900-ms condition, all recognition indices improved similarly for both types of scenes. These findings imply that only when there was time for one fixation on only one picture—which presumably occurred in the 450-ms condition—the content of emotional scenes was processed more accurately than that of neutral scenes. In contrast, this recognition advantage did not occur in the 300-ms condition, presumably because only a very short, incomplete fixation would have been possible before the pictures were masked. This short fixation might have been sufficient to grasp the gist of the emotional scene, hence the higher false alarm rate, but insufficient to identify any specific details. In the 900-ms display condition, no differences occurred as a function of scene valence because there was sufficient time to look at both pictures at least once.

The interpretation of these findings rests on two assumptions. One is related to the time course of overt attention allocation to peripheral stimuli, that is, when an eye fixation starts on a prime picture and how many fixations viewers can make within a specified display duration. This is relevant to whether there is truly peripheral processing of emotional content, or whether foveal vision is indeed required, as argued above. The other assumption is related to bias in attentional orienting, that is, where the first fixation is directed when two pictures are presented simultaneously. We have assumed that, in conditions where there is time for only one fixation on only one picture (i.e., the 300- and the 450-ms displays), an emotional picture would be more likely to receive that fixation than a neutral picture. These assumptions were examined in Experiment 3.

Experiment 3

Pairs of emotional-neutral scenes were presented for 150, 300, 450, or 900 ms, in a within-subjects design, while participants' eye movements were monitored. This allowed us to assess whether a prime picture was foveally fixated during its presentation, which picture of the pair was fixated, when a fixation started, and how many fixations there were on each picture. This is useful to determine (a) the time course of overt attention allocation to peripheral stimuli, (b) whether peripheral emotional scenes are more likely than neutral scenes to capture attention, and (c) whether this can explain the recognition advantage of emotional scenes.

Method

Participants. Twenty-four undergraduates (18 women) participated for course credit.

Stimuli, apparatus, procedure, design, and measures. The same stimuli as in the previous experiments were presented on a 21" monitor, connected to a Pentium IV 3.2-GHz display computer. Participants' eye movements were recorded with an EyeLink II tracker (SR Research Ltd., Mississauga, Ontario, Canada), connected to a Pentium IV 2.8-GHz host computer. The sampling rate of the eyetracker was 500 Hz and the spatial accuracy was better than 0.5°, with a 0.01° resolution in pupil tracking mode. The viewing distance was 60 cm. The location of the prime pictures was readjusted to keep their inner edges 5.2° (5.4 cm) of visual angle away from the central fixation point. The size of the prime pictures was also readjusted accordingly (13.3° by 10.9°; 12.8 by 10 cm).

The procedure and experimental design were identical to those of the previous experiments, with the following exceptions. First, instead of a "flickering" cross as a central fixation point at the beginning of the trial, a drift correction circle generated by the eyetracker software was displayed. The peripheral primes were displayed when the participant was fixating this central circle. Second, each participant was presented with 32 trials in each prime duration condition (i.e., 150, 300, 450, and 900 ms). Trials were randomly assigned to each condition. Of the 32 trials, there were 8 neutral (left) with pleasant (right), 8 pleasant (left) with neutral (right), 8 neutral (left) with unpleasant (right), and 8 unpleasant (left) with neutral (right) pictures. On half of the trials, a target picture was probed; on the other half, a matched picture was probed. For each participant, the stimuli were different in each display condition. This resulted in an experimental design with three within-subjects factors: prime display duration (150 vs. 300 vs. 450 vs. 900 ms), prime valence (pleasant vs. neutral-paired-with-pleasant vs. unpleasant vs. neutral-paired-with-unpleasant), and visual field of the prime (left vs. right).

From the eye movement data during the prime presentation, four measures were informative regarding the attentional assumptions to be tested: (a) the probability that a prime picture received a first fixation within the available display duration indexed selective attention orienting; (b) the time to initiate and perform a saccadic eye movement that landed on a prime picture revealed the speed of orienting; (c) the total number of fixations on each picture indicated attentional engagement; and (d) the total duration of fixations on the picture was used to index the available foveal processing time.

Results

The dependent eye-movement measures were subjected to 4 (prime display duration: 150 vs. 300 vs. 450 vs. 900 ms) × 2 (visual field: left vs. right) × 4 (prime valence: unpleasant vs. neutral-with-unpleasant vs. neutral-with-pleasant vs. pleasant) repeated-measures ANOVAs.

Probability of first fixation. A main effect of display duration, $F(3, 69) = 87.43$, $MSE = 0.062$, $p < .0001$, revealed significant differences between all four levels (ps $< .05$), except between the two longer displays (the means were .001, .262, .354, and .360 for the 150-, 300-, 450-, and 900-ms displays, respectively). The main effects of visual field, $F(1, 23) = 12.58$, $MSE = 0.323$, $p < .01$, and valence, $F(3, 69) = 21.36$, $MSE = 0.032$, $p < .0001$, were qualified by interactions of duration by visual field, $F(3, 69) = 3.60$, $MSE = 0.131$, $p < .025$, and duration by valence, $F(9, 207) = 7.88$, $MSE = 0.014$, $p < .0001$. To decompose these interactions, separate ANOVAs of visual field and valence were conducted for each display condition. The interactions are shown in Figures 3A and 3B. Although there were no significant effects in the 150-ms display, main effects of visual field and valence emerged in the 300-ms display, $F(1, 23) = 10.14$, $MSE = 0.160$, $p < .01$, and $F(3, 69) = 9.34$, $MSE = 0.022$, $p < .0001$; the 450-ms display, $F(1, 23) = 9.00$, $MSE = 0.274$, $p < .01$, and $F(3, 69) = 10.56$, $MSE = 0.020$, $p < .0001$; and the 900-ms display, $F(1, 23) = 4.94$, $MSE = 0.283$, $p < .05$, and $F(3, 69) = 19.06$, $MSE = 0.031$, $p < .0001$, respectively: The probability of fixation was higher for the left than for the right visual field (see Figure 3A), and also higher for the unpleasant and the pleasant prime scenes than for the paired neutral scenes (see Figure 3B; all post hoc contrasts, $p < .05$).

Number of fixations. The main effect of display duration, $F(3, 69) = 111.34$, $MSE = 0.388$, $p < .0001$, reflected a significant difference between all four durations (150 ms = 0.005; 300 ms = 0.292; 450 ms = 0.523; 900 ms = 1.12). The main effect of visual field did not reach statistical significance, $F(1, 23) = 2.41$, $MSE = 0.175$, $p = .13$, but that of valence did, $F(3, 69) = 8.29$, $MSE = 0.224$, $p < .0001$. In addition, there were interactions between duration and visual field, $F(3, 69) = 8.27$, $MSE = 1.306$, $p < .0001$, and between duration and valence, $F(9, 207) = 9.04$, $MSE = 0.185$, $p < .0001$. To decompose these interactions, separate ANOVAs were conducted for each display condition. In the 150-ms condition there were no significant effects due to floor effects (practically no fixations were made). In the 300-ms condition, main effects of visual field, $F(1, 23) = 9.44$, $MSE = 0.207$, $p < .001$, and valence emerged, $F(3, 69) = 12.53$, $MSE = 0.298$, $p < .0001$, which were also significant (or marginally significant) in the 450-ms condition, $F(1, 23) = 3.36$, $MSE = 0.298$, $p = .08$, and, $F(3, 69) = 9.33$, $MSE = 0.034$, $p < .0001$, and in the 900-ms condition, $F(1, 23) = 9.68$, $MSE = 0.142$, $p < .01$, and, $F(3, 69) = 5.41$, $MSE = 0.030$, $p < .01$, respectively. Mean scores are shown in Table 4. More fixations were made on the left than on the right picture in the 300-ms ($M = 0.39$ vs. 0.19) and the 450-ms ($M = 0.59$ vs. 0.45) conditions, with an opposite trend in the 900-ms condition ($M = 1.04$ vs. 1.20). In addition, there were more fixations on the unpleasant and the pleasant pictures than on the neutral pictures both in the 300- and the 450-ms displays; in the 900-ms display, more fixations were made on

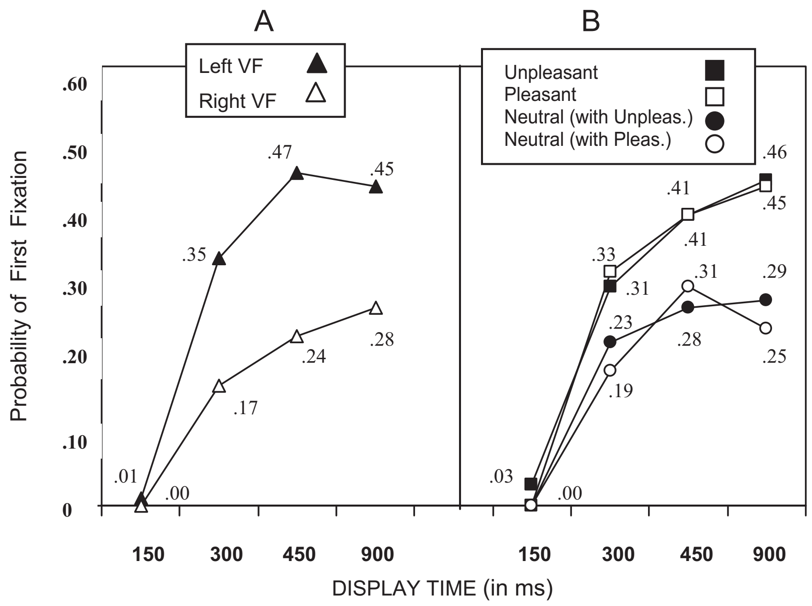

Figure 3. Probability of first fixation as a function of the visual field (Figure 3A) and the emotional valence (Figure 3B) of the prime pictures, for each prime display duration, in Experiment 3.

Table 4
Mean Total Number of Fixations on Prime Picture As a Function of Visual Field, Prime Valence, and Prime Display Duration in Experiment 3

		Neutral with		
Prime duration	Unpleasant	Unpleasant	Pleasant	Pleasant
Left visual field				
150 ms	0.01	0.01	0.02	0.01
300 ms	0.48	0.32	0.29	0.47
450 ms	0.68	0.50	0.55	0.64
900 ms	1.08	1.05	1.07	0.95
Right visual field				
150 ms	0.00	0.00	0.00	0.00
300 ms	0.20	0.15	0.14	0.28
450 ms	0.54	0.38	0.38	0.50
900 ms	1.22	1.20	1.27	1.13

unpleasant than on pleasant pictures ($M = 1.15$ vs. 1.04; all $ps < .05$).

Latency of first fixating a picture and mean duration of fixations. These data could not be properly analyzed statistically due to the minimal or low number of fixations in the 150- and the 300-ms conditions, and the highly unequal number of trials with fixations in the different display conditions. The qualitative analysis is, nevertheless, informative regarding the role of attention in peripheral processing. Mean latency of first fixating a picture was 175, 238, 255, and 258 ms in the 150-, 300-, 450-, and 900-ms conditions, respectively. For those trials in which there was time for a fixation to land on only one picture before it was masked, the mean fixation time was 4, 64, and 188 ms in the 150-, 300-, and 450-ms conditions, respectively. In the 900-ms condition, the fixation time on each of the two pictures was 246 ms.

Discussion

The results of Experiment 3 confirmed our assumptions regarding the time course of deployment of overt attention to the stimuli as a function of prime display duration, and also the assumptions regarding selective attention bias toward emotional stimuli. First, in the 150-ms condition there were practically no fixations on either picture. In the 300-ms condition there was a fixation only on one picture in 53% of trials; moreover, this fixation was of very short duration (it lasted for less than 100 ms prior to picture masking). In the 450-ms condition there was a fixation only on one picture in 71% of trials, and the duration of this fixation approached a typical duration observed in scene viewing studies (nearly 200 ms; see Henderson & Hollingworth, 1998). In the 900-ms condition there was a fixation of typical duration (around 250 ms) on both pictures in 72% of trials, with an equivalent number of fixations made on the emotional and neutral stimuli. Second, emotional scenes were more likely to be fixated than neutral scenes when there was time for only one fixation, that is, in the 300- and 450-ms display conditions. Similarly, when there was time to look at both scenes (in the 900-ms condition), the emotional stimulus was more likely to be fixated before the neutral stimulus, although both stimuli were fixated on most of the trials.

These eye-fixation findings are relevant to explaining the recognition advantage of the peripheral emotional stimuli over neutral stimuli observed in the previous experiments. Such an advantage emerged for the sensitivity scores only when a fixation of normal duration could be made on an emotional picture (i.e., the 450-ms display). In contrast, there was no recognition advantage in the absence of fixations (i.e., 150-ms display), or when the fixation was necessarily very short (i.e., the 300-ms display), or when both the neutral and the emotional picture could be fixated (i.e., the 900-ms display). Accordingly, foveal inspection is required for the identification of specific content of scenes, and allocation of overt attention is needed for the advantage in the processing of peripheral emotional scenes to occur. Nevertheless, the initial selective orienting to emotional scenes suggests that something is seen of the peripherally presented emotional scenes before they are foveally fixated, and that this perception subsequently attracts the eyes to the picture.

General Discussion

We investigated in the present study (a) whether emotional scenes are more likely to be perceived than neutral scenes when they are presented concurrently in the visual periphery, (b) what type of content is processed of the emotional scenes, and (c) the extent to which overt attention is required to recognize specific scene content. A consistent pattern of findings emerged. First, when pairs of prime scenes were presented peripherally for 150 ms, that is, when there was no time to overtly attend to them, the recognition hit rate was at the chance level for all types of scenes, while the false alarm rate was higher for emotional than for neutral scenes. Second, when the display duration of primes allowed one or more fixations to be made on a scene, the first fixation was more likely to be directed to the emotional than to the neutral picture. Third, when there was time to fixate only one of the primes, recognition sensitivity was better for emotional than for neutral scenes, if a fixation of normal duration was possible, that is, the 450-ms display. In contrast, when either only a very short fixation could be made on one picture (i.e., 300-ms display) or there was time to make longer fixations on both pictures (i.e., 900-ms display), there were no recognition sensitivity differences between emotional and neutral scenes. These findings indicate that there is preferential processing of emotional scenes relative to neutral scenes in peripheral vision. However, this peripheral processing yields a coarse impression of the emotional scene (which then governs selective orienting) rather than an advantage in object identification (which requires allocation of overt attention on the scene).

No Advantage in Scene Identification Without Overt Attention

Identification of the specific content of scenes (e.g., the particular people engaged in romantic or violent acts) was assessed by recognition hit rate and the sensitivity index, which assess observers' capability to discriminate between presented and nonpresented— but related—scene content. Three findings reveal that the advantage in the identification of emotional over neutral scenes is dependent on foveal attention, rather than attributable to information acquired by peripheral vision. First, when scenes could not be fixated, the hit rate was at the chance level, and the sensitivity index was equivalent for neutral and emotional scenes. Second, an emotional advantage in discrimination started to show up only when there was enough time to foveally fixate a picture once. Third, correct recognition responses were faster for probes primed by stimuli appearing on the left visual field, which were, in fact, more likely to be fixated than those appearing on the right visual field.

These three findings are consistent with the hypothesis that overt attention to the prime scenes is necessary for the processing of specific semantic information. An alternative interpretation can, nevertheless, be considered: It is possible that increased processing time is sufficient for peripheral identification without foveation. In Experiment 2, we could not separate the contribution of display duration (and therefore processing time) from that of foveation as, with 300- or 450-ms displays, both the probability of fixating the primes and the available processing time increased. One way to discern between these two interpretations would be to separately analyze trials with fixations and those without fixations. This possibility, however, was not viable due to the low or very unequal number of trials with and without fixations. Another way of approaching this issue would be to use a "moving picture" paradigm in which the prime stimuli would move according to the participant's eye position, thus always remaining in the periphery, while varying display time. A related procedure would involve gaze-contingent masking, that is, a 5° moving mask that would block foveal and parafoveal vision of the stimulus while allowing peripheral vision (i.e., beyond the edges of the mask). This technique was, in fact, used by Calvo and Nummenmaa (2007) for parafoveally presented pictures. In their experiments, processing of the affective valence of prime scenes was not obtained the first time a scene was presented for 150 ms, but the priming effect increased across successive presentations of the same scene and became significant after three or four trials. This implies that increased processing time resulting from repeated presentation can be sufficient for affective processing in the absence of foveation. Nevertheless, it must be noted that Calvo and Nummenmaa used affective priming rather than recognition measures, and parafoveal rather than peripheral masking. Accordingly, although their findings suggest that lack of foveation can be compensated for by additional processing time, their conclusions may not be directly generalizable to the current findings.

The current findings extend those of Calvo and Lang (2005), who used the same paradigm and stimuli, but parafoveal—rather than peripheral— stimulus presentation (with the inner edges of the pictures located 2.1° away from the central fixation point). Following a 150-ms prime presentation, Calvo and Lang found that the hit rate for pleasant and unpleasant scenes was above the chance level, whereas the hit rate for neutral scenes was not; moreover, the sensitivity index was higher for emotional than for neutral scenes. Taken together, these two studies reveal that specific semantic information can be obtained from emotional scenes in the parafovea but not in the visual periphery (5.2°). This is consistent with prior research on the processing of nonemotional scenes. Specifically, Henderson, Weeks, and Hollingworth (1999) proposed that no specific semantic content is acquired from scenes in peripheral vision. Rather, semantic processing would begin when the saccade target is foveally fixated, or a fixation has landed relatively near the target region. Other studies using nonemotional scenes as stimuli have provided some support for the proposed distinction between parafoveal and peripheral scene processing. Specific semantic information (i.e., object identity) can be extracted parafoveally from nonemotional scenes (i.e., with target objects located between about 3° and 5° away from the current fixation), but not peripherally (7 to 8°; see Henderson & Hollingworth, 2003, for a review). We have shown that this applies also to emotional scenes.

Gist Processing Advantage in Peripheral Vision

The conclusion that specific semantic information about object identities is not acquired in the visual periphery for either emotional or neutral scenes does not, however, imply that the functional field of view is equally limited for all types of information. There is evidence that gist or category information is grasped with only a very short fixation on a central area of the scene, before details are fixated (see Underwood, 2005), and that this gist then guides object search (de Graef, 2005). In addition, scene categorization (e.g., landscape, traffic scene, etc.) is possible when the scene is presented peripherally (i.e., in the absence of foveal fixations), even though viewers are unable to report precisely what they have seen (Thorpe et al., 2001). Similarly, Gordon (2004) showed that semantic incongruity of a peripheral object in a scene attracts attention to its location, even before there is time to fixate the target object and when viewers are not aware of its identity. This suggests that some type of global processing of scenes occurs in peripheral vision prior to identifying specific objects.

The current study extends these findings by showing that gist information is especially likely to be perceived from peripheral scenes that have emotional content. Emotional significance is related to the gist or global meaning of a scene. In our study, two findings suggest that a global impression of emotional scenes is obtained in peripheral vision prior to landing fixations on the scenes: an initial emotional bias in selective attention orienting (see also Nummenmaa et al., 2006) and an increased false alarm rate for emotional scenes in the 150-ms display condition (see also Calvo, 2006). Regarding initial orienting, the probability of first fixation was higher for emotional pictures than for the paired neutral pictures. As attentional orienting was initiated before a scene was fixated, this finding implies that the viewers had seen something of the emotional scene while it was still in peripheral vision, which then attracted the eyes selectively. The fact that low-level physical characteristics known to attract fixations (e.g., luminance, etc.) were comparable for the emotional and neutral scenes leads us to infer that attentional capture was probably determined by the emotional scene meaning. Nevertheless, peripheral perception was not sufficient to allow the viewer to identify the specific content of emotional stimuli, as revealed by the low hit rate and sensitivity obtained with the shortest prime durations. Rather, the information acquired peripherally must be relatively vague, such as a general impression of the scene (which leads to erroneous stimulus identification, as revealed by false alarm rates), although it is sufficient to subsequently attract overt attention.

Regarding false alarms, when the emotional prime stimuli could not be fixated at all (150-ms display) or could be fixated only very briefly (300-ms display), there was a tendency to misjudge the probe pictures as "presented", even though the probe had not appeared as a prime, but was related in general content and affective valence to a prime picture. False alarms to conceptually similar but visually different probe pictures have been interpreted as an indication of meaningful—although coarse-grained—processing of the prime (Potter, Staub, & O'Connor, 2004). There is, nevertheless, the alternative possibility that the higher false alarm rate could simply be due to a greater similarity in visual features between the target primes and the related probes (which were falsely identified as targets) for the emotional than for the neutral scenes—hence the former being less discriminable than the latter. However, this alternative explanation can be ruled out on three grounds. First, the impact of low-level visual features was similarly minimized for all scene categories by presenting the primes and the probes in different color, size, and spatial orientation. Second, no differences in false alarms emerged between the emotional and the neutral scenes when they were presented foveally or parafoveally (Calvo & Lang, 2005). And, third, in the current study, there was no interaction between valence category and prime-probe relatedness in low-level image properties. This implies that the degree of visual similarity between primes and probes was equivalent for both types of scenes, and leads us to suggest that the effect on false alarm rates under peripheral presentation is indicative of processing of gist, but not specific, information.

Why and How Global Emotional Significance Can Be Processed Peripherally

An important issue is, therefore, the nature of the meaningful gist information that may be extracted in peripheral vision. One candidate is affective valence, that is, whether stimuli are pleasant or unpleasant, good or bad. According to models of automatic emotional processing, the affective content of visual stimuli can be detected preattentively (Bargh, 1997; Zajonc, 2000). One assumption of these models is that the processing of affect does not require detailed perceptual analysis or stimulus identification. Scene valence has indeed been found to be detected from the parafovea in the absence of fixations, under gaze-contingent foveal masking (Calvo & Nummenmaa, 2007). Thus, we argue that peripheral detection of affective content is sufficiently powerful to capture the eyes toward an emotional scene. This notion would account for the results of Experiment 3, as well as those of Calvo and Lang (2004) and Nummenmaa et al. (2006), regarding selective attention orienting. Furthermore, as argued above, such a crude impression is probably responsible for the erroneous probe identifications with short prime durations, as indexed by the results on false alarm rates in Experiments 1 and 2A.

This interpretation is consistent with neurophysiological data. Magnocellular neurons in the visual system are sensitive to peripherally presented stimuli and low spatial frequency (LSF) signals, whereas parvocellular neurons are responsive to foveal stimuli and high spatial frequency (HSF) signals (e.g., Merigan & Maunsell, 1993). Whereas LSF signals convey coarse, global configural information, HSF signals convey fine-grained, detailed visual information. Furthermore, whereas magnocellular channels provide input to subcortical structures such as the amygdala (and then project to the visual cortex), parvocellular pathways project directly to the visual cortex (see Vuilleumier, 2005). Regarding our own findings, this implies that a global impression of scenes could be obtained peripherally through low spatial frequency processing. In addition, given that the amygdala is involved in global or gist emotional processing (see Zald, 2003), it is understandable that the emotional, rather than the neutral, scenes were analyzed by such mechanisms and in such a coarse way. In contrast, precise object identification was not obtained until the scenes were foveally fixated (and thus HSF information could be extracted).

Two types of findings are in agreement with this explanation. First, Vuilleumier, Armony, and Dolan (2003) found that LSF and HSF inputs drove amygdala and visual cortex responses, respectively. Faces with a neutral or a fearful expression were shown centrally for 200 ms. Amygdala responses to fearful relative to neutral expressions were greater for LSF than for HSF faces. In contrast, fusiform cortex responses were greater for HSF than for LSF faces, regardless of expression. Second, Holmes, Green, and Vuilleumier (2005) presented a neutral and a fearful face peripherally for 30 or 100 ms; the faces were replaced by target probes. Faster responses to probes appeared when these replaced an LSF fearful versus a neutral face; in contrast, no response differences appeared between HSF fearful and neutral faces. These findings imply that emotional information can readily be extracted in peripheral vision from low spatial frequency input. An interesting aim for further research is to extend this approach to other types of emotional pictorial stimuli. Our data suggest that also complex scenes and positively valenced stimuli may benefit from low spatial frequency information processing in peripheral vision.

Conclusions

To conclude, (a) there is no semantic processing of specific contents of either emotional or neutral scenes when presented in peripheral vision; nevertheless, (b) global meaning (or an affective impression) may be obtained peripherally from pleasant and unpleasant scenes, which is assumed to be responsible for subsequent (c) bias in attentional orienting to emotional scenes when they are presented simultaneously with neutral scenes; and (d) the processing of specific scene information begins when the scene is overtly attended (i.e., when there is at least one eye fixation on the stimulus). Accordingly, the present study demonstrates that foveal vision and overt attention are necessary to acquire specific scene information, at least for single and short scene displays. There is, however, the possibility that specific peripheral information accrues with additional processing time or repeated presentation of a scene. Nevertheless, covert attention (i.e., in the absence of fixations on the stimulus) can be sufficient to obtain gist information or a general impression of what the peripheral scene is about. As this occurs particularly for emotional scenes, relative to neutral scenes, we conclude that coarse-grained affective meaning is especially accessible to covert attention.

References

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer (Ed.), *The automaticity of everyday life: Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.

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

Calvo, M. G. (2006). Processing of emotional visual scenes outside the focus of spatial attention: The role of eccentricity. *Visual Cognition, 13,* 666–676.

Calvo, M. G., & Esteves, F. (2005). Detection of emotional faces: Low perceptual threshold and wide attentional span. *Visual Cognition, 12,* 13–27.

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

Calvo, M. G., & Lang, P. J. (2005). Parafoveal semantic processing of emotional visual scenes. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 502–519.

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

Center for the Study of Emotion and Attention (CSEA–NIMH). (2005). *The International Affective Picture System: Digitized photographs.* Gainesville: The Center for Research in Psychophysiology, University of Florida.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology, 52,* 95–111.

De Graef, P. (2005). Semantic effects on object selection in real-world scene perception. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 189–211). Oxford, England: Oxford University Press.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11,* 86–89.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45,* 15–31.

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,* 681–700.

Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 760–777.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–293). Amsterdam: Elsevier.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements, visual memory, and scene representation. In M. A. Peterson & G. Rhodes (Eds.), *Analytic and holistic processes in the perception of faces, objects, and scenes* (pp. 356–383). New York: Oxford University Press.

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,* 210–228.

Hermans, D., & Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology, 57,* 97–114.

Holmes, A., Green, S., & Vuilleumier, P. (2005). The involvement of distinct visual channels in rapid attention towards fearful facial expressions. *Cognition and Emotion, 19,* 899–922.

Junghöfer, M., Bradley, M. M., Elbert, T., & Lang, P. J. (2001). Fleeting images: A new look at early emotion discrimination. *Psychophysiology, 38,* 175–178.

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,* 1187–1197.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776.

Lane, R. D., Chua, P., & Dolan, R. J. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia, 37,* 989–997.

Lang, P. J., Bradley, M. M., & Cuthbert, B. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97,* 377–395.

Lang, P. J., Bradley, M. M., & Cuthbert, B. (2005). *International Affective Picture System (IAPS): Technical manual and affective ratings.* Gainesville: The Center for Research in Psychophysiology, University of Florida.

Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways. *Annual Review of Neuroscience, 16,* 369–402.

Mogg, K., Bradley, B., Miles, F., & Dixon, R. (2004). Time course of attentional bias for threat scenes: Testing the vigilance-avoidance hypothesis. *Cognition and Emotion, 18,* 689–700.

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Preferential selective attention to emotional pictures: An eye movement study. *Emotion, 6,* 257–268.

Öhman, A., & Soares, J. J. (1998). Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following non-recognized fear-relevant stimuli. *Journal of Experimental Psychology: General, 127,* 69–82.

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45,* 75–92.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modelling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107–123.

Potter, M. C., Staub, A., O'Connor, D. H. (2004). Pictorial and conceptual representations of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 478–489.

Pourtois, G., Thut, G., Grave de Peralta, R., Michel, C., & Vuilleumier, P. (2005). Two electrophysiological stages of spatial orienting towards fearful faces: Early temporo-parietal activation preceding gain control in extrastriate visual cortex. *NeuroImage, 26,* 149–163.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Sabatinelli, D., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2005). Parallel Amygdala and inferotemporal activation reflect emotional intensity and fear relevance. *Neuroimage, 24,* 1265–1270.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neurosciences, 14,* 869–876.

Underwood, G. (2005). Eye fixations on pictures of natural scenes: Getting the gist and identifying the components. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 163–187). Oxford, England: Oxford University Press.

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

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience, 6,* 624–631.

Wandell, B. A. (1995). *Foundations of vision.* Sunderland, MA: Sinauer.

Zajonc, R. (2000). Feeling and thinking: Closing the debate over the independence of affect. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition. Studies in emotion and social interaction* (Vol. 2, pp. 31–58). New York: Cambridge University Press.

Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews, 41,* 88–123.

Received May 9, 2007
Revision received November 2, 2007
Accepted November 2, 2007 ■