

Enhanced Processing of Emotional Gist in Peripheral Vision

Aída Gutiérrez¹, Lauri Nummenmaa², and Manuel G. Calvo¹

¹Universidad de La Laguna (Spain)

²University of Turku (Finland)

Emotional (pleasant or unpleasant) and neutral scenes were presented foveally (at fixation) or peripherally (5.2° away from fixation) as primes for 150 ms. The prime was followed by a mask and a centrally presented probe scene for recognition. The probe was either identical in specific content (i.e., same people and objects) to the prime, or it was related to the prime in general content and affective valence. The probe was always different from the prime in color, size, and spatial orientation. Results showed an interaction between prime location and emotional valence for the recognition hit rate, but also for the false alarm rate and correct rejection times. There were no differences as a function of emotional valence in the foveal display condition. In contrast, in the peripheral display condition both hit and false alarm rates were higher and correct rejection times were longer for emotional than for neutral scenes. It is concluded that emotional gist, or a coarse affective impression, is extracted from emotional scenes in peripheral vision, which then leads to confuse them with others of related affective valence. The underlying neurophysiological mechanisms are discussed. An alternative explanation based on the physical characteristics of the scene images was ruled out.

Keywords: peripheral vision, emotional, pictorial stimuli, recognition.

En un paradigma de reconocimiento se presentaron fotografías-estímulo (*prime*) de escenas emocionales y neutras durante 150 ms cada una, bien fovealmente (en el centro de fijación visual) o periféricamente (a 5.2° de separación), seguidas por una máscara y una fotografía de prueba (*probe*). La fotografía *prime* y la *probe* podían ser idénticas en contenido específico (las mismas personas y objetos) o únicamente relacionadas en su contenido general y valencia emocional (agradables, desagradables, o neutras). Los resultados mostraron un efecto interactivo de la ubicación espacial y la valencia emocional sobre la tasa de aciertos, pero también la de falsas alarmas y el tiempo de rechazos correctos: No hubo diferencias en estas variables en función de la valencia emocional en la ubicación foveal; en cambio, en la periférica, tanto los aciertos como las falsas alarmas fueron más frecuentes, y el tiempo de los rechazos correctos fue más lento, para las escenas de contenido emocional que las neutras. Los autores concluyen que las personas obtienen una impresión genérica de la valencia afectiva de los estímulos pictóricos en visión periférica, que lleva a confundir las escenas con otras de similar valencia afectiva. Se examinan los mecanismos neurofisiológicos involucrados en este efecto de percepción emocional periférica. Se rechaza la hipótesis de que los efectos del contenido emocional de las imágenes sean debidos a diferencias en las propiedades físicas.

Palabras clave: visión periférica, emoción, estímulos pictóricos, reconocimiento.

This research was supported by Grant PSI2009-07245 from the Spanish Ministerio de Ciencia e Innovación.

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

Emotional stimuli are related to well-being and survival. It is therefore expected that, for adaptive reasons, they have a privileged access to analysis by the cognitive system. To facilitate recruitment of coping resources promptly and speed up responses, the perceptual system must first be biased towards efficiently detecting threat and harm, as well as benefit and pleasure. This implies that emotional stimuli should be preferentially attended when they are competing with non-emotional stimuli, and that the former should be more likely to be perceived at lower temporal thresholds and at more eccentric locations in the visual field. In a sense, as indicated by Vuilleumier (2005), the early sensory processing of stimuli would be enhanced by their affective content. In the current study, we tested two implications of this assumption: Whether emotional scenes are more likely than neutral scenes to be recognized in the periphery of the visual field—relative to the foveal field of vision—and whether this recognition involves an accurate representation of specific features in the scene (objects, people, and actions) or a coarse impression (‘emotional gist’) of the scene emotional valence.

Both neurophysiological and behavioral evidence suggest that emotional pictures (both faces and scenes) are detected more readily than neutral pictures, and also that they are detected earlier, when they are presented *at fixation* in the center of the visual field. Neuroimaging studies have revealed enhanced responses—particularly in the amygdala, but also in occipital and parietal regions of the cortex—to emotional stimuli relative to neutral stimuli (e.g., Sabatinelli, Bradley, Fitzsimmons, & Lang, 2005). Studies assessing event-related potentials in the brain (ERPs) have shown amplified responses to emotional visual stimuli involving early sensory components (e.g., P1 and N1 at 120–150 ms) as well as later components (after 300–400 ms) (e.g., Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000). Similarly, a wide range of behavioral measures (e.g., Calvo & Avero, 2006; Hermans, Spruyt, De Houwer, & Eelen, 2003, with an affective priming paradigm) and peripheral physiological responses (e.g., Öhman & Soares, 1998, with electrodermal assessment) have shown enhanced processing of emotional stimuli when they are presented at fixation very briefly, subliminally, or masked.

Some studies have assessed whether emotional scenes have a privileged access to the cognitive system even when they are presented in *peripheral* areas of the visual field and outside the focus of overt attention, i.e., without eye fixations. Calvo (2006) and Calvo, Nummenmaa, and Hyönä (2008) presented emotional-neutral pairs of visual scenes peripherally (with their inner edges 5.2° away from fixation) as primes for 150 ms, followed by a 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, if no fixations on the primes were allowed, the false alarm rate—but not the hit rate or sensitivity (A')—was higher for emotional than for neutral

scenes. The authors argued that the sensory enhancement for emotional stimuli in peripheral vision involves a coarse impression that is extracted from them, rather than an accurate processing of specific features. The reasoning is as follows. False alarms represent incorrect “yes” responses to probe pictures that had *not* been presented as primes, but that were *related* in *emotional valence* to the presented primes. If participants obtained only a global impression (i.e., the affective gist) of the emotional primes, this would lead them to confuse a probe with a non-presented, but related prime, and to wrongly accept the emotional probe as presented. If, instead, the specific features of the emotional prime stimuli had been processed, this should have been reflected in an increased hit rate *and* sensitivity for the emotional probes, which was not the case.

In the current study, we will compare this explanation—called the *emotional gist* hypothesis—with an alternative explanation—called the *formal similarity* hypothesis. According to the emotional gist hypothesis, a coarse impression of the affective content of emotional scenes is obtained in peripheral vision, and false alarms reveal such emotional gist processing. In visual cognition studies, “scene gist” has been defined as a semantic representation of the global properties of a scene, or the semantic category or abstract concept of the situation depicted in the scene (e.g., like “seeing the forest without representing the trees”: Greene & Oliva, 2009; see Castelano & Henderson, 2008; Rousselet, Joubert, & Fabre-Thorpe, 2005). “Emotional scene gist” would be a particular case of the more general scene gist. It refers to the affective valence of a scene, with different levels of generality, i.e., whether the scene is emotional or not, whether it is pleasant or unpleasant, or even the category of pleasantness (e.g., erotic) or unpleasantness (e.g., attack). Scene gist processing has been investigated in previous research using recognition paradigms by assessing false alarms (e.g., Castelano & Henderson, 2008; Greene & Oliva, 2009; Potter, Staub, & O’Connor, 2004): Incorrect “yes” responses to probe scenes that are conceptually related (but different in specific features) to prime scenes are thought to indicate that the gist of the prime scene was processed, and led to confusion between presented primes and (non-presented, but globally consistent) probes. Accordingly, the greater false alarm rate for emotional pictures in peripheral vision would indicate that a coarse impression of their affective valence is extracted in the recognition paradigm used by Calvo (2006) and Calvo et al. (2008).

In contrast, according to the *formal similarity* hypothesis, the higher false alarm rate could simply be due to visual similarity—regardless of meaning—between the target primes and the related probe scenes (which were falsely identified as targets) that were used in the Calvo (2006) and Calvo et al. (2008) studies. This type of visual similarity could be greater for the emotional than for the neutral scenes; hence the former would be less discriminable than the latter

and thus produce more false alarms. This would, actually, imply that emotional significance would *not* be processed peripherally, or, at least, with no advantage over non-emotional content. A number of potential, purely formal aspects of the pictures could be involved. Among them, the role of low-level visual characteristics, such as luminance, contrast, energy, and color saturation were ruled out by Calvo et al. (2008). Using the same stimuli and design as in the current study, emotional and neutral scenes did not differ in any of these physical properties. Nevertheless, there are other formal similarity factors such as the shapes of the people or objects in the scenes, the relative figure/ground area, the number of elements in the scene, the type of depicted activity, whether the people are directly gazing at the viewer or not, etc. A greater false alarm rate would be produced for emotional than for neutral scenes when the probe is presented if there is more similarity in these formal aspects between the prime and the probe for emotional than for neutral scenes.

To examine this alternative, formal similarity hypothesis in the current study, we used two different approaches. First, we directly assessed some additional physical properties of the emotional and the neutral images in three ways: (a) we computed pixel-by-pixel correlations of the intensities (i.e., grayscale luminance) of the corresponding target and matched stimuli; (b) we used principal component analyses (PCA) and assessed how much of the intensity variation of each image pair could be explained by the first PC—the more variation the first PC explains, the more similar the images are; and (c) in a recent study (Nummenmaa, Hyönä, & Calvo, 2009), we assessed the perceptual saliency of the images using the Neuromorphic Vision Toolkit (iNVT) developed by Itti and Koch (2000; Itti, 2006), where perceptual saliency is computed by combining local contrast, spatial orientation of features in the scene, and image energy. Evidence supporting the formal similarity hypothesis should involve greater pixel-by-pixel correlations, more similarity, and higher saliency for the emotional than for the neutral scenes. A lack of differences between neutral and emotional scenes in these measures would rule out this hypothesis.

In a second approach to test the formal similarity hypothesis, we presented the emotional and the neutral prime scenes both at fixation (i.e., foveally; in the center of vision) and peripherally, followed by probe scenes in a recognition paradigm. The number of potential formal differences between pictures is unknown, as well as the relative role of each formal difference, and so it may not be sufficient to try to isolate and assess each, and make the emotional

and the neutral categories comparable (as in the first approach). Rather, a more comprehensive control condition involves presenting the same pictures in a foveal and in a peripheral condition. The formal properties of the scenes are the same in both conditions. Accordingly, if the greater false alarm rate (and longer reaction times for false alarms) for emotional than for neutral scenes in the peripheral condition were due to those formal properties, the same pattern of false alarms would occur *also* in the foveal condition. In contrast, the formal similarity hypothesis would be ruled out if the false alarm rate is greater for emotional than for neutral scenes *only* in the peripheral condition, and there is a greater increase in false alarms in the peripheral (vs. the foveal) condition for emotional than for neutral scenes.

We used essentially the same paradigm as Calvo et al. (2008) and Calvo (2006) did, with two major differences. First, a foveal display condition was used to present the prime scenes, in addition to the peripheral condition. Second, in the peripheral condition, instead of presenting two simultaneous prime scenes—one emotional and one neutral on each trial—we presented only *one* prime scene—unpleasant, pleasant, or neutral—either to the left or to the right of fixation simultaneously with a meaningless picture (a random combination of colors) that was equated with the prime picture in luminance. In the peripheral condition, only one prime scene was presented on each trial to make this condition comparable with the foveal condition regarding the number of meaningful stimuli. Furthermore, in the peripheral condition, a simultaneous meaningless picture was presented to prevent attentional orienting to the prime scene due to its abrupt appearance. The equivalence in luminance between the two simultaneous stimuli was aimed at cancelling out the automatic attraction of overt attention that would have otherwise produced a singleton prime scene. With this manipulation, the prime scene was made available (foveal condition) or unavailable (peripheral condition) to overt attention. The peripheral prime scene (and the meaningless picture) was presented with a visual angle of 5.2° away from a central fixation point for a short duration (150 ms) to prevent any fixations.¹ Following the prime picture and a backward mask, there was a recognition test, where we used two types of probe pictures. The probe was either *identical* in content to the prime scene (target probe, used in target-present trials), to assess hits and the acquisition of specific information; or the probe was *related* in affective valence (matched probe, used in target-absent or catch trials), to assess false alarms and gist information.

¹ A 150-ms display at a 5.2° distance from the fixation point has been found to prevent fixations on the peripheral picture. In Calvo et al. (2008), the mean latency of the first saccade to the picture was 175 ms (hence above the 150-ms display duration), the probability that the peripheral picture was fixated (within the 150-ms display) was less than 1%, and the mean fixation time in these few cases was only 4 ms. This implies that, in the conditions of the current study, the peripheral pictures were also very unlikely to be fixated.

Method

Participants

Forty-eight (36 female) psychology undergraduates (between 18 and 25 years of age) participated for course credit. Half of them were randomly assigned either to the foveal or the peripheral presentation condition.

Stimuli and Stimulus Characteristics

We used 128 picture stimuli. For the *target-present* trials, 64 digitized color photographs were presented as target stimuli, of which 32 were neutral in affective valence (i.e., non-emotional), 16 were unpleasant and 16 were pleasant. All target pictures portrayed people, either (a) in a variety of daily activities (neutral scenes), or (b) suffering threat or 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 which was matched with one of the target pictures in topic, composition, presence of people, valence, and arousal. Accordingly, the target and the matched scenes were similar in general content and emotionality, but their specific content and details were different. Out of the total sample of 128 experimental pictures, 87.5% of them were selected from the International Affective Picture System (IAPS; Center for the Study of Emotion and Attention, 2005); 12.5% were obtained from other sources.¹

Valence and arousal ratings for each picture were analyzed in a Valence category (unpleasant vs. neutral vs. pleasant picture) by Relatedness (target vs. matched pictures) ANOVAs (see Calvo et al., 2008). There were significant differences between all three categories on valence, $F(2, 122) = 811.15, p < .0001$, and arousal scores were higher for the unpleasant and the pleasant scenes than for the neutral scenes, $F(2, 122) = 32.28, p < .0001$, with no arousal differences between the pleasant and unpleasant stimuli. The effects of relatedness and the interaction were not significant (both $F_s < 1$).

Using Adobe Photoshop, 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 Calvo et al., 2008). The luminance and contrast of some of the original pictures were readjusted to make the three valence categories comparable. The one-way (valence: unpleasant vs. neutral vs. pleasant) ANOVAs showed no significant differences in any of these image characteristics (all $p_s > .25$; for energy: $F(2, 61) = 3.13, p \geq .082, ns$).

By means of the iNVT algorithm (Itti & Koch, 2000), a stimulus-driven saliency map was computed for each visual scene (see Nummenmaa et al., 2009). Briefly, the visual input is first decomposed and processed by feature (e.g., local contrast, orientation, and energy) detectors mimicking the response properties of retinal neurons, lateral geniculate nucleus, thalamus, and V1. These features are then integrated for a neural saliency map, which is a graded representation of the visual conspicuity of each pixel in the image. Salient areas (or objects) thus stand out from the background, including other surrounding objects. One-way (valence) ANOVAs showed no significant differences in the maximal saliencies or the mean saliencies between the unpleasant, neutral, and emotional scenes (all $p_s > .10$).

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° by 10.9° . When presented for recognition as probes, all pictures were in grayscale, in enlarged size (35.8° by 26.9°), and their left-right orientation was mirror-reversed. This change in formal properties (size, color, and spatial orientation) from prime to probes was made to reduce the contribution of purely physical factors to recognition (i.e., that the participants could not rely simply on these formal cues to compare the prime and the probe), and thus increase the contribution of semantic processing to recognition (i.e., by requiring the participants to identify the prime-probe similarities in content beyond formal differences). The original colors were used for the prime scenes because color has been found to contribute to gist processing (Castelhano & Henderson, 2008; Goffaux et al., 2005). We, nevertheless, assumed that, once obtained

² 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).

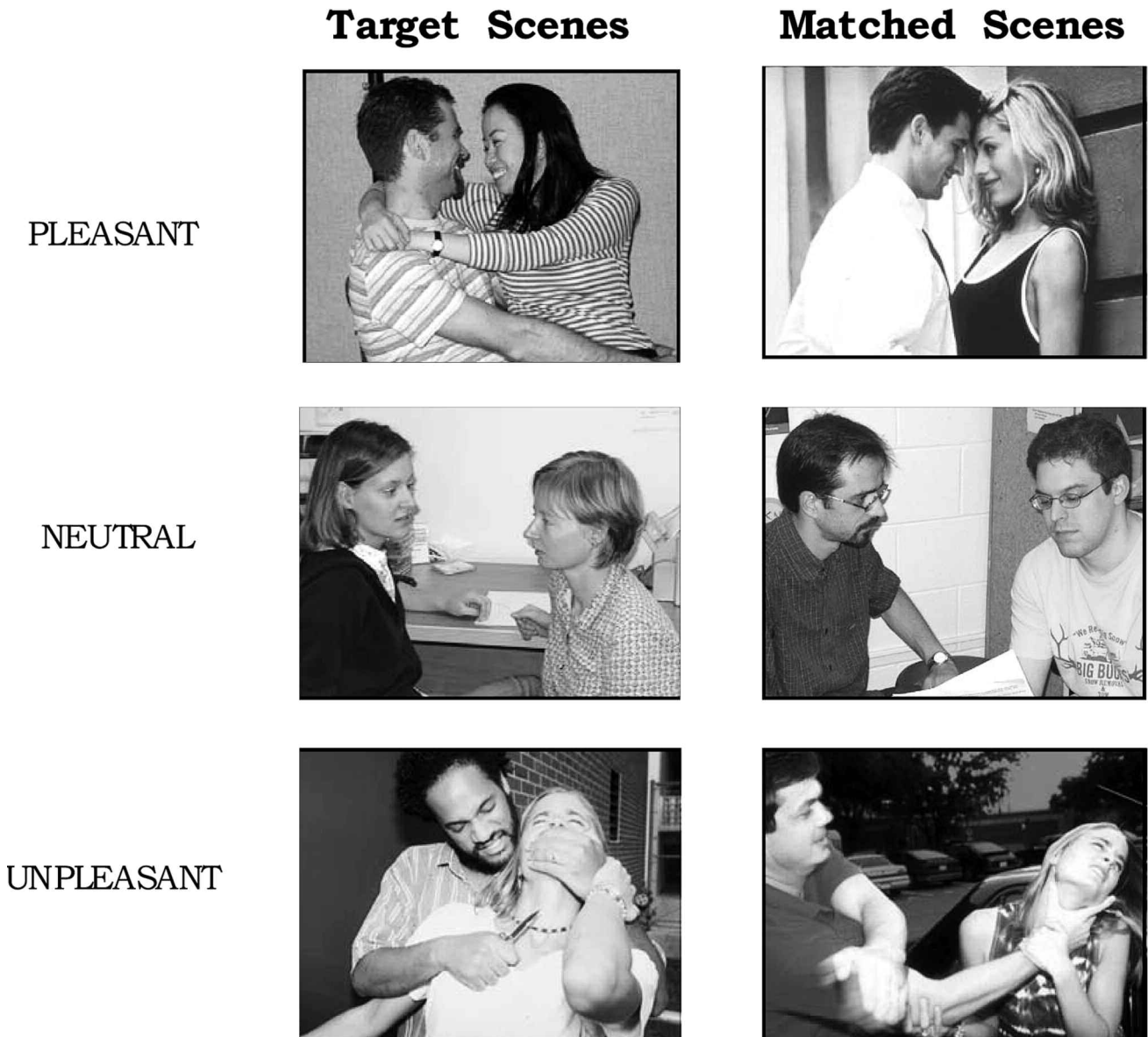


Figure 1. Examples of target and matched (related in content and affect) neutral and emotional scenes.

the gist from the prime picture, the transformation into a grayscale probe picture would not affect the recognition of emotional and the neutral pictures differently.

Apparatus and Procedure

The pictures were displayed on a SVGA 17" monitor with a 100-Hz refresh rate, connected to a Pentium-IV computer. The E-Prime software 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 a prime photograph, and that they should fixate at a central cross (in the peripheral condition; or at the center of the screen, in the foveal condition), until a probe photograph appeared for recognition. It was made clear that the photograph for recognition would increase in size, change from color to grayscale, and change in left-right orientation, with respect to the prime picture. The participant was to respond as quickly as possible whether this formally modified probe photograph had, nevertheless, the same content ("the same people doing the same things") as the prime, by pressing a "Yes" key (L) or a "No" key (D). Sixteen practice trials were followed by 128 experimental trials, in two blocks.

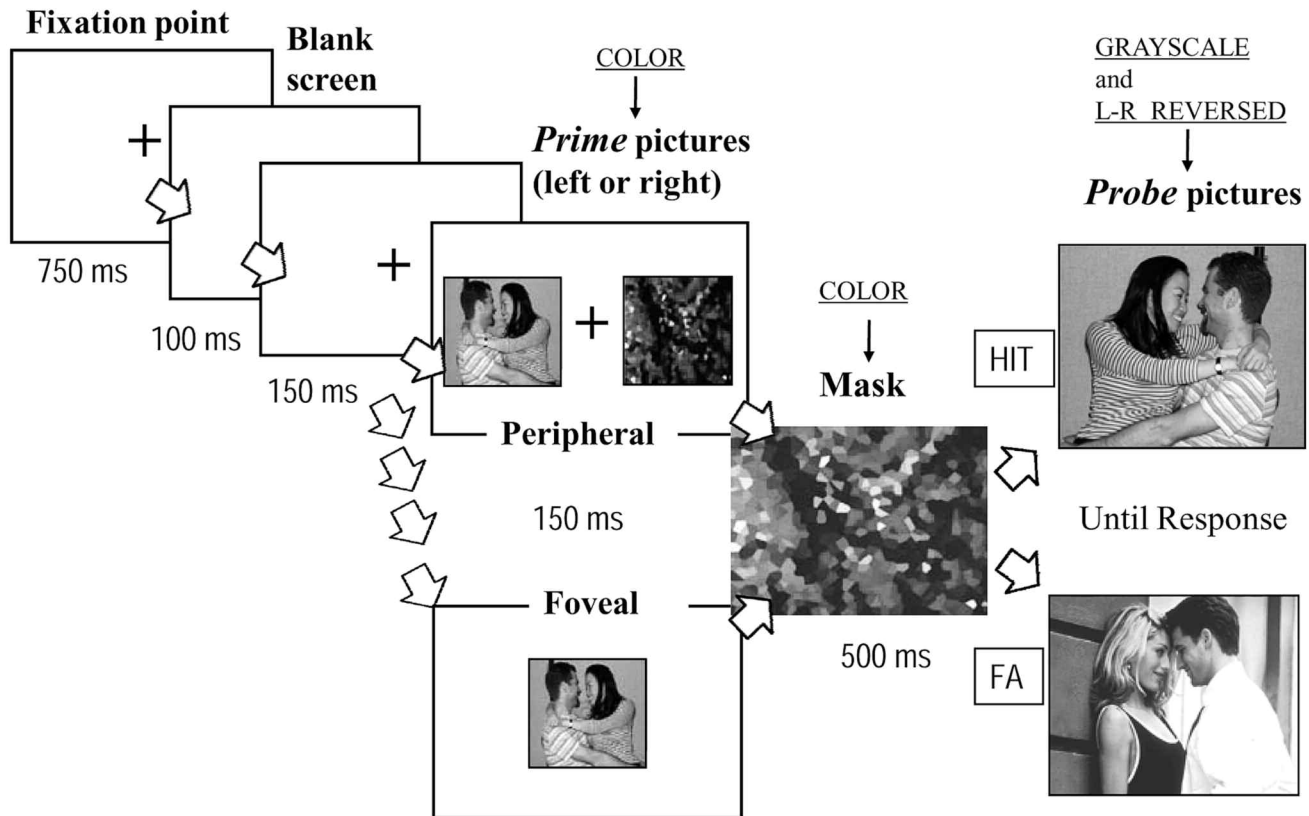


Figure 2. Sequence of events within a trial. Only one probe picture appeared on each trial, either the emotional or neutral target scene (hit), or the emotional or neutral matched scene (false alarm). In the peripheral display condition, both the content scene (people) and the meaningless (color) picture appeared; in the foveal display condition, only the content scene appeared (at the center of the screen).

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 blinking of the cross served to capture the viewer's attention on the central location just before the prime pictures appeared and prevent saccades to the peripheral scene. Next, in the peripheral condition, the prime scene and a meaningless picture (a random combination of colors; yet equivalent in luminance) was displayed peripherally for 150 ms. In the foveal condition, only the prime picture appeared for the same duration. In the peripheral condition, the prime scene (and also the simultaneous meaningless picture) appeared laterally, to the left or right of a central fixation cross. The inner edge of both these pictures was located horizontally at 5.2° of visual angle (4.33 cm) from the central fixation cross. In the foveal condition, the size of the prime picture was the same as in peripheral condition, although the foveal picture appeared at fixation, in the center of the visual field, and there was no simultaneous meaningless picture. 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.

Design

The experiment involved a 2 (Location: foveal vs. peripheral) \times 3 (Emotional valence of the picture: unpleasant vs. neutral vs. pleasant) \times 2 (Prime-probe relationship: identical vs. related in content) \times 2 (Visual field of the prime: left vs. right) factorial design. Location was a between-subjects factor; the others were within-subjects factors.

Each participant saw each picture twice as a prime (once followed by an identical target probe and once by a matched, related probe, in different blocks), and once as a probe in the recognition test. In each block, 50% were target trials and 50% were matched trials. The target probe appeared in one block (either first or second) and the corresponding matched probe appeared in the other block (second or first). On target trials, the probe was identical in specific content, although different in color, size, and orientation to the prime. On catch trials, the probe was related in topic and emotionality to the prime, although different in form (color, size, and orientation) and specific content. In the foveal condition, there were two counterbalancing conditions, depending on whether a prime was followed by an identical probe or a related probe in the first or the second block.

In the peripheral condition, there were four counterbalancing conditions, depending on whether a prime appeared on the left or right visual field, and whether it appeared followed by an identical or a related probe in the first or the second block. The trials were presented in random order within each block.

Results

Recognition Performance

The probability of hits (PH; i.e., correct “yes” responses to probes on target-present trials) and false alarms (PFA; i.e., incorrect “yes” responses to probes on target-absent trials) in recognition performance were converted to an A' index of sensitivity or discrimination (see Snodgrass & Corwin, 1988); where $A' = 0.5 + (PH - PFA) * (1 + PH - PFA) / (4 * PH) * (1 - PFA)$. A' scores vary from low to high sensitivity on a scale ranging from 0 to 1.

The mean scores for each dependent variable are presented in Table 1. A series of mixed 2 (Location) \times 3 (Emotional Valence) \times 2 (Visual Field) ANOVAs were conducted for the probability of hits and false alarms, sensitivity, and reaction times for hits and for correct rejections. We used Bonferroni-corrected post hoc comparisons ($p < .05$) for all multiple comparisons involving emotional valence. In the absence of visual field effects, this factor was collapsed.

For the probability of hits, there were main effects of location, $F(1, 46) = 50.77, p < .0001$, and valence, $F(2, 92) = 4.85, p < .025$, as well as an interaction, $F(2, 92) = 3.25, p < .05$. For false alarms, the analysis yielded also main effects of location, $F(1, 46) = 10.45, p < .01$, and an interaction, $F(2, 92) = 4.82, p < .025$. Reaction times for hits were affected only by location, $F(1, 46) = 23.81, p < .0001$, whereas reaction times for correct-rejection responses were affected by location, $F(1, 46) = 21.88, p < .0001$,

valence, $F(2, 92) = 5.62, p < .01$, and their interaction, $F(2, 92) = 3.36, p < .05$. Finally, a main effect of location emerged on the sensitivity (A') index, $(1, 46) = 46.43, p < .0001$. Consistently for all dependent variables, the strong effect of location indicated that the hit rate and sensitivity were higher, the false alarm rate was lower, and reaction times were shorter in the foveal than in the peripheral display condition.

The most interesting results were concerned with the interactions of location and emotional valence. To decompose the interactions, one-way (emotional valence) ANOVAs were conducted separately for the foveal and the peripheral condition for each dependent variable. In the foveal condition, there were no significant differences in any dependent variable as a function of emotional valence (all F s < 1). In contrast, in the peripheral condition, the hit rate was higher for both unpleasant and pleasant scenes than for neutral scenes, $F(2, 46) = 5.80, p < .01$, and the false alarm rate was also higher for both unpleasant and pleasant scenes than for neutral scenes, $F(2, 46) = 4.84, p < .025$. Interestingly, and consistently with the false alarm results, in the peripheral condition, reaction times for correct rejection responses to both unpleasant and pleasant probes were slower than to neutral scenes, $F(2, 46) = 6.62, p < .01$. Furthermore, in the same vein, correct-rejection responses (i.e., “no” responses to probes non-presented as primes) were slower than hit responses (i.e., “yes” responses to probes presented as primes) for emotional scenes, M unpleasant = 999 vs. 907 ms, $t(23) = 2.07, p < .05$; and M pleasant = 986 vs. 903 ms, $t(23) = 2.39, p < .025$ but not for neutral scenes ($M = 910$ vs. 871 ms, $p = .41, ns$).

Low-level Image Similarity Between the Target and the Matched Scenes

As an additional test of the formal similarity hypothesis (see the introduction), we assessed the visual similarity of

Table 1

Mean Probability of Hits and False Alarms (FAs), Sensitivity Scores (A'), and Reaction Times (RT, in ms) for Hits and for Correct Rejections (CR), as a Function of Emotional Valence of Scenes, in the Foveal and the Peripheral Presentation Conditions

Valence	Hits	FAs	A'	RT Hits	RT CR
Foveal Display					
Unpleasant	.899	.097	.943	660	738
Neutral	.895	.107	.941	643	726
Pleasant	.913	.097	.947	664	735
Peripheral Display					
Unpleasant	.766 ^a	.251 ^a	.841	907	999 ^a
Neutral	.705 ^b	.195 ^b	.834	871	910 ^b
Pleasant	.780 ^a	.257 ^a	.841	903	986 ^a

Note. A different superscript ^{a b} letter indicates significant differences between valence categories.

the target and the matched unpleasant, neutral, and pleasant stimuli by two complementary approaches. First, we computed pixel-by-pixel correlations of the intensities (i.e., grayscale luminance) of the corresponding target and matched stimuli. Second, we used principal component analyses (PCA) and assessed how much of the intensity variation of each image pair could be explained by the first PC—the more variation the first PC explains, the more similar the images are. Mean correlations (M unpleasant = .10 vs. M neutral = .05 vs. M pleasant = .14) and variations explained by the first PC (M unpleasant = 62.65 vs. M neutral = 64.65 vs. M pleasant = 65.64) were subjected to one-way ANOVAs. These demonstrated that there were no statistically significant differences in the correlation, $F = 1.26$, $p = .29$, or PCA, $F < 1$, $p = .68$, based similarities of the unpleasant, neutral, and pleasant stimulus pairs. Accordingly, any confusion between the presented (target) and non-presented (matched) probes in the recognition of emotional scenes could not be attributed to these being more physically similar than the corresponding target-matched neutral scenes.

Discussion

When prime scenes were presented peripherally for 150 ms, the probe recognition hit rate—but also the false alarm rate—was higher for emotional than for neutral scenes. As a consequence, the A' index was equivalent for emotional and neutral scenes. As the sensitivity index represents the discrimination between presented and non-presented information, these findings imply that no more information about the specific content of scenes (i.e., the *identity* of the portrayed objects, people, or actions) was acquired from the emotional than from the neutral scenes. Rather, the false alarm data suggest that the tendency of more hits for emotional pictures is simply contaminated by their susceptibility to confounds with related, but non-presented, information. The reaction times for correct rejections were consistent with this interpretation. It took longer to say “no” to probes that had not been presented as primes when they were emotional than when they were neutral. Furthermore, correct rejection responses were slower than hit responses for emotional scenes, but not for neutral scenes. This suggests that participants had a tendency to say “yes” to non-presented emotional items, which gave rise to more false alarms and/or caused a delay in correct rejection responses (due to the additional time needed for inhibiting the wrong “yes” tendency).

Calvo et al. (2008) proposed that this enhanced false alarm rate and the delay in correct rejections is caused by confusion between presented and non-presented scenes, and that such confusion comes from the affective similarity between the emotional primes and probes. More specifically, the emotional valence of the prime scenes would be

assessed to some extent when these are presented peripherally. Valence processing would involve only a coarse or vague impression about the prime scene. When a probe related in valence appears, the viewer would compare or match the representation he or she has extracted from the prime with the content of the probe. If little specific information about the identities of the objects in the scene is perceived peripherally, a tendency to admit that the prime and the probe are the same would be likely to occur when there is strong relationship in affective valence between them. In other words, if emotional valence—but not specific contents—is perceived peripherally, the viewer will rely on the affective similarity between the prime and the probe to respond in the recognition test. In contrast, valence would not be present or prominent in the neutral scenes and therefore they would be less prone to induce this kind of response bias. Consistent with this interpretation, false alarms to conceptually similar but visually different probe pictures have been interpreted as an indication of meaningful, albeit coarse, processing of the prime, i.e., scene gist (Castelhano & Henderson, 2008; Greene & Oliva, 2009; Potter, Staub, & O'Connor, 2004). In the current study, the meaningful information susceptible to confusion would involve the affective similarity between otherwise physically different primes and probes. This is interpreted as emotional scene gist.

The lack of differences between neutral and emotional scenes in the foveal condition is contrary to the alternative, formal-similarity interpretation of the false alarm and correct rejection time data. According to the formal-similarity hypothesis, the differences in false alarms between the emotional and the neutral scenes in the peripheral condition could be due to a greater similarity in visual features (e.g., shapes, colors, etc.) between the target primes (i.e., presented) and the related probes (i.e., non-presented as primes) for the emotional than for the neutral scenes. A greater formal similarity would make the scenes less discriminable and thus prone to more false alarms. Against this explanation, however, the foveal condition clearly indicated that the emotional and the neutral scenes were similarly recognizable, with no differences in any dependent variable, including false alarms and correct rejection times. In the foveal condition, recognition performance increased in all aspects, including accuracy and speed. This probably reflects the fact that specific information about object identities was obtained from all types scenes to a greater extent than in the peripheral condition. In the foveal condition, the role of gist meaning or a coarse impression provided by valence decreases because the viewer can rely on direct evidence from the details of the scene. In line with this, the correlation and PCA-based similarity estimates, as well as analysis of first and second-order image statistics, showed that the unpleasant, neutral and pleasant target-matched scene pairs were equally visually similar. Consistently, there were no differences between the emotional and the neutral pictures

in color saturation, luminance, RMS contrast, or energy (Calvo et al., 2008) or in perceptual saliency (Nummenmaa et al., 2009). Hence it cannot be argued that the pleasant and the unpleasant target-matched pairs were visually more similar than the respective neutral pairs, or that this accounted for the false alarm and correct rejection time effects.

Accordingly, the current data refine the view that emotional content enhances perception of stimuli (Vuilleumier, 2005). They support the hypothesis that a global impression or gist, rather than object identification or genuine specific semantic processing of emotional scenes, is obtained in peripheral vision, outside the focus of overt attention. Recent studies have provided increasing support for extrafoveal processing of non-emotional semantic content (Underwood, 2005). The gist or scene category can be extracted in the visual periphery, but gist perception does not necessitate identification of individual objects in the scene (Gordon, 2004; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bühlhoff, 2001). In general, our findings are consistent with the view that gist and specific scene features can be encoded separately (see also Sampanes, Tseng, & Bridgeman, 2008). Emotional significance is related to the gist or global meaning of a scene; it is not a single feature of the scene, but rather a global impression of its pleasantness or unpleasantness. In the current study, we have shown that this meaningful gist of emotional valence can be obtained in peripheral vision. Eye-movement studies have also provided findings consistent with this interpretation. When emotional and neutral scenes are presented simultaneously, the first fixation is directed more likely to the emotional than to the neutral scene, even when the scenes are equivalent in low-level image properties (e.g., luminance, etc.) known to affect eye movements at early stages (Calvo et al., 2008; Nummenmaa, Hyönä, & Calvo, 2009). The preferential early orienting to the emotional pictures in peripheral vision suggests that “something” of these scenes is seen prior to fixation, which then attracts attention. The information acquired peripherally must be relatively vague, such as a general impression of the scene, hence leading to erroneous (i.e., false alarms) or delayed (i.e., correct-rejection reaction times) stimulus identification, although it is sufficient to subsequently attract overt attention, if there is time for saccades.

A final issue is concerned with *why* the emotional gist of visual scenes is processed in peripheral vision. One reason is that the emotional scene content can be conveyed by low spatial frequencies that are accessible to the magnocellular

neurons receiving inputs from the peripheral retina (in contrast, the specific features of the scenes are conveyed by high spatial frequencies that are accessible only under foveal presentation, through parvocellular channels). This argument is backed up by recent findings indicating that, although progressive low-pass spatial frequency filtering (i.e., high spatial frequencies are eliminated) dampens subjective valence and arousal ratings of complex emotional pictures, such filtering does not fully ‘neutralize’ the valence and arousal of the scenes (De Cesare & Codispoti, 2008). Furthermore, a growing number of data reveal a close functional relationship between the extraction of low frequencies and emotional processing of both faces (see Eimer & Holmes, 2007) and more complex scenes (Carretié, Hinojosa, López-Martín, & Tapia, 2007). The magnocellular layer in the lateral geniculate nucleus (LGN) projects both to primary visual cortex (V1) and the amygdala. Given that the amygdala is involved in rapid global or gist emotional processing (see Zald, 2003) and that it responds more vigorously to low- than to high-pass filtered fearful faces (Vuilleumier, Armony, Driver, & Dolan, 2003), it has been proposed (Vuilleumier & Pourtois, 2007) that the extrageniculostriate pathway would be crucially involved in emotional processing. Within this background, it is understandable that, in the current study, the emotional, rather than the neutral, scenes were analyzed in such a coarse way in peripheral vision: An enhanced global impression of the emotional scenes could be obtained through low spatial frequency processing, which would be quickly undertaken by the amygdala.³

References

- 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., & Avero, P. (2006). Affective priming with pictures of emotional scenes: The role of perceptual similarity and category relatedness. *Spanish Journal of Psychology*, *9*, 10-18.
- Calvo, M.G., Nummenmaa, L., & Hyönä, J. (2008). Emotional scenes in peripheral vision: Selective orienting and gist processing, but not content identification. *Emotion*, *8*, 68-80.
- Carretié, L., Hinojosa, J. A., López-Martín, S., & Tapia, M. (2007). An electrophysiological study on the interaction between emotional content and spatial frequency of visual stimuli. *Neuropsychologia*, *45*, 1187-1195.

³ Nevertheless, this does not rule out the possibility of cortical processing of emotion being involved too, as the magnocellular layer in the LGN also projects to V1. Intracranial recordings have actually demonstrated that both the amygdala (Oya, Kawasaki, Howard, & Adolphs, 2002) and the ventral prefrontal cortex (Kawasaki et al., 2001) discriminate between unpleasant and neutral scenes within the same time window around 120 ms, suggesting parallel subcortical and cortical emotion processing. Accordingly, if the emotional visual information is conveyed by the low spatial frequencies, it can be relayed rapidly to both cortical and subcortical networks involved in emotional processing, which guarantees effective emotion detection under various viewing conditions such as when the target is away from fixation.

- Castelhano, M.S., & Henderson, J.M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 660-675.
- Center for the Study of Emotion and Attention [CSEA-NIMH] (2005). *The International Affective Picture System: Digitized photographs*. Gainesville, FL: 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 Cesarei, A., & Codispoti, M. (2008). Fuzzy picture processing: Effects of size reduction and blurring on emotional processing. *Emotion*, *8*, 352-363.
- Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, *45*, 15-31.
- Greene, M.R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*, 137-176.
- Gordon, R.D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 760-777.
- Gofaux, V., Jacques, C., Moraux, A., Oliva, A., Schyns, P.G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, *12*, 878-892.
- Hermans, D., & Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology*, *57*, 97-114.
- Itti, L. (2006). Quantitative modeling of perceptual saliency at human eye position. *Visual Cognition*, *14*, 959-984.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489-1506.
- Kawasaki, H., Adolphs, R., Kaufman, O., Damasio, H., Damasio, A. R., Granner, M., Bakken, H., Hori, T., & Howard III, M.A. (2001). Single-neuron responses to emotional visual stimuli recorded in human ventral prefrontal cortex. *Nature Neuroscience*, *4*, 15-16.
- Nummenmaa, L., Hyönä, J., & Calvo, M.G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, *6*, 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*, 305-323.
- 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.
- Oya, H., Kawasaki, H., Howard, M.A., III, & Adolphs, R. (2002). Electrophysiological responses in the human amygdala discriminate emotion categories of complex visual stimuli. *Journal of Neuroscience*, *22*, 9502-9512.
- 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.
- Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition*, *12*, 852-877.
- 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.
- Sampanes, A.C., Tseng, P., & Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Research*, *48*, 2275-2283.
- 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 Neuroscience*, *14*, 869-76.
- 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: Oxford University Press.
- 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.
- Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585-594.
- Vuilleumier, P., & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia*, *45*, 174-194.
- Zald, D.H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews*, *41*, 88-123.

Received September 29, 2008

Revision received January 22, 2009

Accepted February 22, 2009