Emotions are often object related—they are about someone or something in the world. It is yet an open question whether emotions and the associated perceptual contents that they refer to are processed by different parts of the brain or whether the brain regions that mediate emotions are also involved in the processing of the associated content they refer to. Using functional magnetic resonance imaging, we showed that simply combining music (rich in emotion but poor in information about the concrete world) with neutral films (poor in emotionality but rich in real-world details) yields increased activity in the amygdala, hippocampus, and lateral prefrontal regions. In contrast, emotional music on its own did not elicit a differential response in these regions. The finding that the amygdala, the heart of the emotional brain, responds increasingly to an emotional stimulus when it is associated with realistic scenes supports a fundamental role for concrete real-world content in emotional processing.

Keywords: amygdala, emotion, fMRI, hippocampus, prefrontal cortex

Introduction

Emotions have evolved as evaluations of events in the world that guide adaptive behavior (Lazarus 1991; Clore and Ortony 2000). In order to serve this function, they must be closely linked to information about the concrete real world (i.e., be grounded). For example, when being attacked by a bear, for an appropriate emotion (i.e., fear of the bear) and response (i.e., running away) to emerge, the evoked emotion must refer to the perceived representation of the bear and the danger it presents. When looking into the brain from this perspective, it is reasonable to expect that a neural mechanism that mediates an emotional state will also be involved in processing aspects of its associated content. It has been demonstrated that areas traditionally thought to be involved solely in cognitive processes, such as working memory and perception, can be modulated by emotion (Rotshtein et al. 2001; Perlstein et al. 2002). However, the sensitivity of emotion-related brain regions to associated perceptual content has not been sufficiently considered.

In order to study the relative involvement of a given brain region in processing emotions and associated content, one needs to manipulate them independently. To achieve that, in the present study, 12 healthy subjects passively watched short film clips and listened to short music clips in a cross-modal design while their brain activity was measured by functional magnetic resonance imaging (fMRI). Subjects were presented with music clips—rich in emotion but poor in concrete content; film clips—poor in emotion but rich in concrete real-world details; and music and film clips in combination. Music clips were of negative, positive, or neutral emotional character, whereas film clips were all emotionally neutral. Thus, when neutral film and music clips appeared together, the neutral film acquired the emotional valence provided by the music (Fig. 1).

Emotions elicited by music are relatively abstract, as music provides no representation of the immediate real world. On the other hand, when music is embedded in a film, the cinematic scene, rich in details and objects, grounds the abstract musical emotion in the real world and gives it a concrete meaning (Cohen 2001). As was shown before, music cannot only set the emotional tone of a film but also radically affect the way in which the film is interpreted and remembered (Lipscomb and Kendall 1994; Boltz 2001). Thus, in our paradigm, the emotion elicited by the combination of film and music differed from the film alone in its emotionality and from the music alone in its added concrete content. Accordingly, we assumed that brain regions that process an emotion together with the concrete content it refers to would be preferentially activated by the emotional combination compared with the separate film and music.

It is an open question whether such an integrated process recruits a unique neural mechanism or involves dedicated neural systems that handle each process alone, such as amygdala for emotion and prefrontal cortex (PFC) and visual cortex for associated perceptual content. The amygdala comprises the heart of the brain’s emotional system. It might be specifically relevant to our question because in a rodent model of fear conditioning, it was shown to be critical for tying emotional responses to external stimuli (LeDoux 2000). Based on primate studies, it was proposed that the amygdala is acting as a modulator to ensure that emotional responses are appropriate to the external stimuli and to the social context (Bachevalier and Malkova 2006). Moreover, together with the hippocampus, the amygdala is thought to facilitate the association of emotions to spatial and episodic information about the world (Phelps 2004). Altogether, it seems that the amygdala serves a comparable role in various emotional tasks differing greatly in complexity and explicit cognitive load, such as simple fear conditioning, contextual conditioning, recognition of fearful facial expressions, and emotion-guided decision making (Bechara et al. 1999; Adolphs et al. 2005). However, because the activity in the amygdala was not directly compared between these different tasks, it left open the possibility that the associated content by itself modulates its function. It has been shown that an attention-based task decreases amygdalar response to unrelated emotional stimuli (Phan et al. 2002; Pessoa et al. 2005). However, in these studies, the perceptual tasks were deliberately unrelated to the emotional processing and were therefore in competition with it for neural resources. To the best of our knowledge, the significance of perceptual content that constitutes a part of an emotional
processing, rather than interferes with it, has not yet been assessed. The question remains whether the amygdala mediates not only the affective aspect of emotional processing but also its associated content. In our study, differential amygdala response to music alone and to its combination with movies would support the latter.

The PFC is thought to play an important role in behavioral control, mediating perception and action (Fuster 2001), particularly when guided by internal states, such as affect and motivation (Miller and Cohen 2001). Miller and Cohen also suggested that the PFC achieves this function by maintaining patterns of activity relating to both visual perception and emotional evaluations among other processes. Specifically, the medial PFC has been implicated in emotion-guided decisions and conflict processing, whereas the lateral PFC has been implicated in guiding and shaping perceptual decisions (Bechara et al. 1999; Miller and Cohen 2001). Interestingly, a functional connection between the emotionally associated medial PFC and the cognition-associated lateral PFC has been suggested before, based on their widespread anatomical connections in primates (Barbas 2000).

Consistent with such an integrative role, the PFC has widespread connections to other brain regions, including the amygdala, the hippocampus, high-order visual and auditory cortices, and the lateral temporal cortex (Fuster 2001). It was therefore reasonable to expect prefrontal regions to also be involved in binding abstract emotional states (as evoked by the music clips) with perceived representations of the world (as presented in the movie clips).

**Materials and Methods**

**Subjects**

Fourteen healthy, right-handed volunteers (aged 23–30 years, 6 females) participated in the imaging experiment. All signed an informed consent form that was approved by the Tel Aviv Sourasky Medical Center and Tel Aviv University ethical committees. Subjects with previous knowledge about the experiment were not recruited. Two subjects (1 female) were excluded from the final analysis due to excessive head movements during the scan. Fifteen subjects (aged 20–29 years, 6 females) participated in the behavioral emotional judgments of the stimuli. Nine female subjects (aged 25–38 years) participated in the behavioral concrete content judgments of the stimuli. No subjects participated in both the behavioral and the imaging studies.

**Visual and Auditory Stimuli**

The visual stimuli consisted of 12 film clips of 12 s each. The film clips were selected from commercial films according to the following requirements: 1) film clips were of emotionally neutral character; 2) coupled with emotional music, film clips could be interpreted as both negative and positive; 3) film clips contained no dialogue; and 4) film clips portrayed no widely familiar actors or scenes. The auditory stimuli consisted of 12 music clips, 12 s each. The music clips contained no human voices or sounds related to specific objects (e.g., a clock ticking, footsteps, etc.). The music clips were of positive (joyful), negative (scary), or neutral (simple and monotonic) emotional tones (4 clips of each type). The neutral music was used to control for the effect of nonemotional auditory stimuli. Positive music clips were selected from commercial popular music. Two negative clips were composed (cut and edited) from a horror film sound track and from a classical music piece (using WaveLab 4.0 from Steinberg, Steinberg Media Technologies, GmbH, Hamburg, Germany). Two negative clips and all the neutral clips were created in-house using music software (Cubase.
VST 5 from Steinberg, Reaktor 3.0 and Kontakt 1.0 from Native Instruments Software Synthesis, GmbH, Berlin, Germany). The average decibel level of all music clips was made equal using Wavelab 4.0 from Steinberg.

Experimental Paradigm
Each subject was presented with 12 emotionally neutral film clips and 12 music clips of negative, positive, or neutral valence. The stimuli were presented as 12-s epochs separated by 9-s blank epochs (gray screen). Arrangement of the stimuli was done with film-editing software (Adobe Premiere 5) and then burned on a DVD disc, which was used for both the behavioral and imaging experiments.

Within-subject arrangement: film clips (F), music clips (M), and their combination (FM) were order balanced in the 6 combinatorial possibilities of presentation (F FM M; FM FM; M FM etc.). The clips of a specific "trio" (F1 FM1 M1) were distanced from each other as far as possible (average distance = 1.1875 epochs per 3.941 min, standard deviation 1.32 epochs per 0.662 min). To avoid priming effects, all emotionally neutral scenes (silent film, neutral music, and neutral combination) were equally positioned after negative and positive conditions. An additional film clip combined with calm music was positioned as the first stimulus, which was ignored in the data analysis.

Across-subject arrangement: three versions of the paradigm were created, each presented to a third of the subjects (see Supplementary Table 1). Please note that although all versions consisted of the exact same set of stimuli, the specific film–music coupling is different per each group of subjects. This way, a specific film was not associated with a specific valence type of music across subjects. For example, a film clip that was combined with positive music in one version was combined with negative music in the second and with neutral music in the third. This design was used in order to avoid any bias due to specific film features.

Stimulus Judgment Experiment
A preceding behavioral study was performed in a quiet room using the same experimental paradigm as in the imaging experiment. Subjects viewed the same experimental paradigm on a TV screen using a DVD player. As customary in studies of emotion, subjects were requested to rate each stimulus on 2 scales (Lang et al. 1993). One was a "valence" scale, in which the valence of the emotion elicited by each clip was rated on a 10-level bipolar scale (−5, "very negative" to +5, "very positive"). Subjects were instructed using these examples: "an example of negative emotion is fear and an example of positive emotion is joy." The other scale was a "arousal" scale, in which the arousal of the emotion elicited by each clip was rated on a unipolar scale of 10 points (0, "not arousing" to 10, "very arousing"). Film and music clips that were not rated as expected separately or in combination were discarded, and alternative clips were used instead. All clips included in the experiment were rated by 15 subjects. A different group of 9 subjects rated to what extent the to 10, "very arousing"). Film and music clips that were not rated as

Table 1. Please note that although all versions consisted of the exact same set of stimuli, the specific film–music coupling is different per each group of subjects. This way, a specific film was not associated with a specific valence type of music across subjects. For example, a film clip that was combined with positive music in one version was combined with negative music in the second and with neutral music in the third. This design was used in order to avoid any bias due to specific film features.

Resolution and 2 mm slice thickness (field of view [FOV] = 24 × 24, matrix: 256 × 256, time repetition/time echo [TR/TE] = 20/3 ms), was acquired for each subject. This anatomical study allowed for volume statistical analyses of signal changes during the experiment. Functional $T_2^*$-weighted (TR/TE/flip angle = 3000/50/90; with FOV 24 × 24 cm$^2$, matrix size 128 × 128) images were acquired (24 oblique slices, thickness: 4 mm; gap 1 mm, covering the whole cerebrum) in runs of 6288 images (262 images per slice). An oblique slice, individually determined based on the occipital end and orbital-frontal surface of each brain, was used in order to maximize the signal from limbic structures.

Functional Imaging Data Analysis
Imaging data were analyzed using BrainVoyager 2000 software package (Brain Innovation, Maastricht, The Netherlands). Raw functional images were superimposed and incorporated into the 3D data sets through trilinear interpolation. The complete data set was transformed into Talairach space (Talairach and Tournoux 1988). Preprocessing of functional scans included head movement assessment (scans with head movement >1.5 mm were rejected) and correction of high-frequency temporal filtering by the removal of linear trends. To allow for $T_2^*$-equilibration effects, the first 6 images of each functional scan were rejected. The 3D group statistical parametric maps were calculated using a general linear model (GLM), in which all stimuli conditions were defined as positive predictors. The GLM consisted of 9 conditions: 3 modality types (film, music, and combination) × 5 musical valence types (negative, positive, and neutral). A uniform lag of 1 repetition was used to account for the hemodynamic response shift.

Region of Interest Analysis
Region of interest (ROI) analysis was applied to the PFC, the amygdala, and the hippocampus. In addition, visual and auditory regions were also examined, including the lateral occipital complex (LOC), the parahippocampal place area (PPA), the fusiform face area (FFA), and the planum temporale (PT) (the rationale for choosing these specific regions is elaborated in Results). The ROIs were defined based on a conjunction of functional and anatomical criteria: 1) Only voxels activated by a contrast of all "combination" conditions versus baseline in a multivariate GLM (n = 12, P < 0.005, fixed effects model, uncorrected) were included. This contrast marks the network of regions generally involved in processing visual and/or auditory stimuli. As negative, positive, and neutral conditions are all included in this contrast, it holds no bias in terms of emotion. 2) Anatomical boundaries were defined in Talairach space for each ROI. The PFC, the amygdala, and the hippocampus were marked using the Talairach atlas (Talairach and Tournoux 1988). The PFC was defined as Brodmann areas 9, 10, 11, 12, 25, 32, 45, 46, and 47. Special care was taken to avoid regions that are known to be involved in premotor or visuomotor functions (Brodmann areas 6, 8, and 44) (Faw 2003). The LOC was marked as a box-shaped region encompassing Talairach coordinates reported by Grill-Spector et al. (1999). The PPA and FFA were marked as 10 × 10 × 10 voxels box-shaped regions centered around Talairach coordinates reported by Epstein et al. (2003) and Kanwisher et al. (1997), respectively. The PT was marked according to Talairach coordinates reported by Westbury et al. (1999). The central Talairach coordinates and the number of voxels of each ROI are reported in Supplementary Table 2 online. ROIs that were combined together were averaged according to their relative number of voxels.

For each ROI, time courses were collected from all subjects. In each time course, average percent signal change was calculated for each epoch relative to implicit baseline (the average signal during all blank epochs) using Microsoft Excel. Two-way analysis of variance, with modality (combination, film, and music) and specific film–combination–music triplets (triplet 1, triplet 2, triplet 3, and triplet 4) as factors, was separated to apply each type of emotion using STATISTICA software. Thus, a separate "combination" effect was anticipated between the combined and each of the separate conditions. A second comparison was attempted between the combination and the summation of the separate condition.

Correlation Analysis
Correlation of ROI time courses was examined between the amygdala and anterior hippocampus during the separate and combined film and
music conditions. For each subject, all readings of the blood oxygenation level-dependent (BOLD) signal were collected for each condition in each of the ROIs, and then the data sets from the amygdala and anterior hippocampus ROIs were correlated with linear regression analysis using Microsoft Excel for each condition separately. The degree of correlation was then compared between the different conditions on the group level using a paired t test (n = 12).

**Isolation of a Combination-Specific Network**

First, a conjunction analysis of [combination > film] and [combination > music] was performed (n = 12, random effects analysis, P < 0.002 uncorrected, extent threshold = 2 voxels per 1.65 mm³). Then, the resulting gray matter clusters were masked by an inclusive mask obtained by the contrast [combination > baseline] and by exclusive masks obtained separately by the contrasts [music > baseline] and [film > baseline] (n = 12, random effects analysis, P < 0.05 uncorrected, extent threshold = 2 voxels per 1.65 mm³). This 2-step analysis identified clusters of activation that were active exclusively in response to the combination (and not to the separate conditions) and whose activity was significantly higher in response to the combination compared with the separate conditions. This same procedure was performed for the negative, positive, and neutral conditions separately.

**Results**

**Behavioral Experiment**

In a preceding study, 15 subjects (different from the fMRI subjects) rated the emotion elicited by the various clips in terms of valence and arousal. Figure 2 shows that all film clips and neutral music clips were rated as relatively nonemotional (nonarousing and neutral), whereas the negative and positive music clips were rated as highly arousing and clearly negative or positive. Most importantly, the combinations of music and film were rated similarly to the same music clips on their own.

The various clips were also rated in terms of their concrete content in an additional behavioral study. The film clips and the combinations were similarly rated as very concrete, whereas the music was rated as significantly less concrete (Fig. 2G–I). Thus, as was the case in Boltz’s study (Boltz 2001), it was the music that determined the emotional tone of the combination, whereas the films provided the music with concrete content.

**fMRI Experiment**

**Emotion and Content at the Limbic Level**

The amygdala and anterior hippocampus were examined with a ROI approach (see Materials and Methods). The ROIs were anatomically marked out of a general functional cluster that was obtained by the contrast of all the combination conditions versus baseline (Fig. 3C). This contrast marks the network of regions generally involved in processing visual and/or auditory stimuli and holds no bias in terms of emotion. As discussed above, the amygdala and anterior hippocampus are known to closely cooperate on various aspects of emotional processing (LeDoux 2000; Phelps 2004) and when collected separately indeed showed a similar pattern of results. The results of the amygdala and anterior hippocampus ROIs are therefore described side by side.

The averaged percent signal change was obtained for each region and analyzed within each valence type. The rationale behind this approach is to compare “physically identical” stimuli
within each subject (see Materials and Methods for more details and Supplemental note 1 for further discussion). In both the amygdala and anterior hippocampus, response was greater during the negative combinations than during both films and negative music clips separately (Fig. 3A, B). This effect is also evident in statistical parametric maps obtained for the separate and combined film and negative music (Fig. 4). The positive conditions elicited a similar effect only in the amygdala ROI (Fig. 3A). In contrast, for the neutral control condition, both the amygdala and hippocampus did not show additional activity in response to the combination compared with the film clips alone. Thus, additional activation was obtained only when the music provided an emotional theme to the film and not simply because of salient simultaneous processing of music and film. Moreover, this occurs despite the fact that neither the amygdala nor the anterior hippocampus was differentially responsive to emotional music by itself (see Fig. 3). As verified in the preceding behavioral study (Fig. 2), the combination of film and emotional music did not differ from the same music alone in the emotional properties of valence and arousal, assuring that the addition of film did not affect “emotionality”; rather, it added “concrete content.” Hence, it is demonstrated that manipulating associated content and emotions can modulate both amygdalar and hippocampal activities. For an additional analysis confirming that the selective response to the combined film and music conditions does not stem from a different response to the film-alone condition, see Supplemental note 2. A further comparison between the combination and the summation of the separate conditions was attempted and yielded no statistically significant results. It is presumed that the lack of significance is due to

Figure 3. Limbic ROIs. Averaged activations during the combined and separate music and film in the amygdala (A) and anterior hippocampus (B). The ROIs in the right and left hemispheres did not significantly differ in the contrast of interest and are presented together. Error bars represent standard error of mean. *P < 0.05 and ns, nonsignificant; n = 12. The amygdala and the anterior hippocampus ROIs are delineated on an axial section (C). The ROIs were anatomically marked out of a general functional cluster that was obtained by the contrast of all the combination conditions versus baseline (P < 0.005, uncorrected, fixed effect) (see Materials and Methods).
great signal variance in the indicated ROIs, but it may also reflect
the neural mechanism of integration, which is not necessarily
a linear summation of the separate conditions.

We then asked whether the mere combination of film and
music affects not only the level of activity in the amygdala and
hippocampus but also the synchronization between them. The
time courses of activation of the amygdala and anterior
hippocampus ROIs (Fig. 3C) showed a statistically significant
correlation during all conditions. However, a borderline signif-
icant increase in correlation ($P < 0.055$) was found between
the right amygdala and hippocampus when the film and negative
music were combined compared with the correlation during
the separate film and music (Fig. 5). In contrast, no increase in
correlation was found during the neutral and positive combi-
nations compared with the separate film and music. These
results suggest that the functional association of the amygdala
and the hippocampus was selectively strengthened when
a negative emotion was integrated with concrete content.

Emotion and Content at the Prefrontal Level
The PFC ROI was defined as all voxels within the PFC that were
activated by all the conditions of combined films and music (see
Materials and Methods). The PFC ROI encompassed some
heterogeneous areas in the lateral aspect of the PFC, including
Brodmann areas 9, 45, 46, and 47 (Fig. 6B). Despite its
recognized role in emotional processing, no significant activa-
tions were obtained in the medial PFC (Brodmann areas 10, 11,
12, 25, and 32). However, this negative finding should be
interpreted with caution because relatively low BOLD signals
were obtained in most subjects from medial prefrontal areas
that may indicate on susceptibility artifacts. The fact that the
tasks in our paradigm were all implicit may also contribute to

Figure 4. Activity in response to the combined and separate conditions. Activations obtained for the film condition (A), negative combination condition (B), and negative music condition (C) compared with baseline ($P < 0.002$, uncorrected, random effects analysis, $n = 12$). Talairach coordinates $x = 43, y = −13, z = −17$. 
the diminished activation of the medial PFC, a region known for its part in explicit tasks such as appraisal (Bechara et al. 1999).

Averaged activation in the lateral PFC ROI demonstrated enhanced activity in response to the negative and positive combination conditions in comparison with the film and music alone (Fig. 6A). This effect is also evident in statistical parametric maps obtained for the separate and combined film and negative music (Fig. 4). In contrast, no additional activity was found in response to the neutral combination in comparison with the film or music on their own. Hence, it was the combination of a concrete visual scene with "emotional" music that preferentially activated the lateral PFC. These results suggest that the lateral PFC is involved in associating concrete content about the world with abstract emotions.

A Neural Network that Integrates Emotions and Concrete Content

We next set out to isolate, with no a priori hypothesis, a network of regions that selectively mediate emotions with concrete content (elicited by the combination of music and films) but are not involved in processing abstract emotions or concrete content by themselves (elicited by the separate music and films, respectively). A 2-step whole-brain analysis first identified clusters of activation whose activity was significantly higher in response to the combination than in response to the separate conditions. Then, using a masking procedure, only those clusters that responded exclusively to the combination (and not to the separate conditions) were isolated (see Materials and Methods).

For the negative conditions, such combination-specific clusters were obtained in a number of regions including the right amygdala, right hippocampus, middle and inferior frontal gyri, and lateral temporal regions (Table 1, negative combination effect). For the positive conditions, combination-specific clusters were seen in similar frontal and lateral temporal areas (Table 1, positive combination effect). Combination-specific clusters were much fewer for the neutral conditions and in areas distinct from those found for the negative and positive conditions (Table 1, neutral combination effect). Thus, it seems that this distinct network of regions, which includes our a priori hypothesized ROIs, participates in processing a property that is unique to the combined emotional music and film and not present in each alone, namely, an emotion with a concrete real world–based content.

Figure 5. Time course correlation between the right amygdala and hippocampus ROIs. Linear correlation between the BOLD signal of the right amygdala and anterior hippocampus during the separate and combined conditions. Averaged correlation coefficients were compared between the separate and combined conditions, with a paired t-test. Resulting with a t value of 1.73 between film and combo coefficients and a t value of 2.0 between music and combo coefficients. Error bars represent standard error of mean. *P < 0.055, n = 12.

Figure 6. Prefrontal ROI. Averaged activations during the combined and separate music and films (A). The ROIs in the right and left hemispheres did not differ in the contrast of interest and are presented together. Error bars represent standard error of mean. *P < 0.05, **P < 0.001, ns, nonsignificant, n = 12. The prefrontal ROI is delineated on sagittal and axial sections. (B) The ROIs were anatomically marked out of a general functional cluster that was obtained by the contrast of all combination conditions versus baseline (P < 0.005, uncorrected, fixed effect).
Emotion and Content at the Sensory Level

A linkage between emotions and representations of the external world may begin already at the sensory level of processing. Indeed, previous imaging studies have shown modulation of activity in visual regions by the emotional valence of visual stimuli, and such modulation was shown to be mediated by the amygdala (Rotshtein et al. 2001; Phan et al. 2002; Vuilleumier et al. 2004). However, most of these studies have not manipulated emotions independently of their visual content.

We examined whether high-order sensory areas are also involved in the association of abstract emotions with concrete scenes by ROI analysis that was applied to high-order visual areas. Areas that were previously shown to be modulated by emotion include the LOC, the FFA, and the PPA. The PT, which is thought to constitute the secondary auditory cortex (Shapleske et al. 1999), was analyzed as the auditory ROI. As expected, the high-order visual areas showed a preference for visual stimuli but no additional activation when music was added to the film. No significant difference was found between the different visual ROIs, and so only results from the LOC ROI are presented (Fig. 7A). As a mirror image, the PT showed a preference for auditory stimuli, with equal or even diminished activation when film was added (Fig. 7B). Thus, no additional activity was found in response to the combination of music and film in the sensory ROIs.

To examine whether this result was produced due to our specific choice of sensory regions, we performed a whole-brain analysis using the contrast of [positive combination > negative combination > same films without music] in all the regions that were activated by films on their own ($P < 0.05$, uncorrected, random effects analysis). This contrast isolates all visually responsive clusters that demonstrated increased activity in response to the addition of emotional music. No significant activations were found at a threshold of $P < 0.002$ (uncorrected, random effects analysis). Thus, our data do not demonstrate emotional modulation of high-order visual areas. This is not surprising because we did not manipulate the emotional content of the visual stimuli; we manipulated emotions independently of their visual content.

Discussion

Using a novel cross-modal approach with emotional music and neutral film clips, we were able to demonstrate selective brain responses to musical emotions depending on the presence of concrete contents conveyed by films. In both ROI and whole-brain analyses, the amygdala, the anterior hippocampus, and the lateral PFC exhibited additional activity in response to the combination of negative music and neutral film clips compared with the same clips presented separately. Interestingly, these regions, including the amygdala, were not activated selectively by emotionality in music when presented without a film. These findings strongly suggest that the brain exerts a preferential response to emotional stimuli when these are associated with concrete content. The cognitive approach in affective science defines emotions as evaluations of events in the world that guide adaptive behavior. It thus predicts that emotions are processed together with the contextual representations they refer to (Lazarus 1991). The present findings demonstrate that even in the amygdala, a core region of the limbic system, emotional processing is closely linked to its associated content.

Integrative Emotional Processing in Limbic Areas: Associated Content and Valence

In line with numerous neuroimaging studies (Blood et al. 1999; Levitin and Menon 2003; Schmitzhorst and Holland 2003), music in the present study elicited no activity above baseline in the amygdala when presented on its own. A few studies have shown increased or decreased amygdalar response to emotional music (Koelsch et al. 2006; Blood and Zatorre 2001, respectively). However, in these studies, the selective responses were elicited by comparing types of music and not by contrasting the unpleasant condition to baseline and may therefore reflect a differential amygdalar response to the music itself rather than sensitivity to emotionality in music. Our results may also appear to conflict with a study by Gosselin et al. (2005), in which patients with medial temporal lesions were impaired in emotional recognition of fear elicited by music presented alone, leading to the conclusion that the amygdala plays a role in the recognition of danger in a musical context. However, although the impairment was attributed to a damaged amygdala, the lesions that were studied consistently encompassed extensive temporal and paralimbic regions, including the parahippocampal gyrus, an area previously implicated in emotional processing of music (Blood et al. 1999).

Strikingly, in our study, the same emotional music clips that did not elicit an amygdalar response when presented on their own did evoke a significant additional effect in the amygdala when combined with neutral film clips. To explain this finding, one might suggest that the amygdala is more sensitive to emotional stimuli associated with visual input (Phan et al. 2002). This claim corresponds with evidence in primates of widespread anatomical projections from the amygdala to virtually all levels of the visual cortex (Amaral et al. 1992) as
well as with numerous neuroimaging studies in humans showing close relations between activation in the amygdala and high-order visual areas (e.g., Rotshtein et al. 2001; Bleich-Cohen et al. 2006). On the other hand, imaging studies showed that the human amygdala was activated during anticipation of outcome following risky choices in a computerized game, regardless of the external stimuli (Kahn et al. 2002). Moreover, extensive evidence from animal lesion studies supports the role of the amygdala in processing affective auditory information in a conditioning context (LeDoux 2000). Similarly, several human imaging studies have shown amygdalar activations in response to affective auditory stimuli such as unpleasant words, laughter, and crying (Sander and Scheich 2001; Maddock et al. 2003; Sander et al. 2003). Altogether, these various data make the possibility that the amygdala only processes information conveyed by external visual modality very unlikely.

In light of these previous studies, our results point to another possibility whereby amygdala’s activation may be driven by the emotion-related perceptual content of a stimulus that determines the goal of action in the world. In all the aforementioned studies that report amygdalar activation, the emotion studied by various modalities had a “concrete and immediate meaning,” related to approach or withdrawal behavior. A fearful face, a fearful voice, a conditioned stimulus, and a risky choice: these can all signal concrete and immediate danger; a recalled emotion or unfamiliar music alone cannot. The amygdala is a relatively primitive neural structure that has undergone little change throughout mammalian evolution (LeDoux 2000). It thus makes sense that its role may be specific for processing emotions that have a clear adaptive survival value. Accordingly, in the present experiment, an emotion that refers to real-world content (elicited by the combination of music and film) yielded

---

**Figure 7.** Sensory ROIs. Averaged activation during the combined and separate music and films in the LOC (A) and PT (C) ROIs. The right and left ROIs did not significantly differ in the contrast of interest and are presented together. Error bars represent standard error of mean. *P < 0.05, **P < 0.001, ns, nonsignificant; n = 12. The LOC (B) and PT (D) ROIs are delineated on sagittal and axial sections (see Materials and Methods). The ROIs were all anatomically marked based on its own reported Talairach coordinates.
amygdalar response, even though emotionality was conveyed through an abstract music. Finally, Seifritz et al. (2003) showed that the sound of infants crying does not elicit amygdalar activation in nonparents, but does so in parents, and so the authors claimed that parenthood causes neuroplastic changes in the brain. We suggest that parenthood alters the interpretation of the sound of infants crying and gives it concrete and immediate subjective emotional meaning. Similarly, in a recent neuroimaging study, it was shown that the amygdala was more active with closed than open eyes in the dark while subjects were listening to the same emotional music (Lerner et al. 2005). This finding suggests that concrete and immediate emotional meaning is not necessarily dependent on external cues but rather can be elicited by focusing on the self-oriented world. That said, it has to be mentioned that our ROI analysis was limited to those regions of the amygdala that reacted to the combination of music and film. It is quite possible that these regions primarily carry out an integrative processing role, whereas other regions in the amygdala have a more confined, abstract emotional processing role.

Like the amygdala, the anterior hippocampus also demonstrated increased activations to negative emotional music when combined with neutral films in comparison with emotional music alone. Whereas the amygdala has been implicated in emotional processing (LeDoux 2000), the hippocampus has been implicated in processing spatial representations and episodic memories (Burgess et al. 2002). Also, it seems that both the amygdala and hippocampus have access to conceptual information about the world, as a large proportion of their neurons selectively responds to category-specific stimuli as shown by single-cell recordings in humans (Kreiman et al. 2000). Moreover, the amygdala and the hippocampus were suggested to play complementary roles in binding emotional memories to contextual cues in fear conditioning (Phelps 2004). Our results extend previous findings in showing that even when emotionality remains unchanged, emotion-related episodic information by itself (provided by the film) modulates amygdalar activity. Furthermore, without a concrete visual scene, an abstract musical emotion elicited no amygdalar response, suggesting that reference to concrete meaning may be necessary for amygdalar emotional processing. Finally, correlation of right amygdalar and hippocampal activities was significantly increased when film and negative music were combined. We thus propose that grounding of an abstract emotion in the concrete world primarily involves the amygdala, the hippocampus, and the interaction between them. We cannot ignore a possible ROI overlap between those 2 anatomically neighboring regions. That said, it has to be remembered that the increase in correlation was found only for the combined condition of the negative emotional valence. If the correlation was a result of overlaps, we would not have seen such valence-dependent differences in correlation levels.

The results of the present study regarding the amygdala and the hippocampus proved to be more consistent for negative emotions than for positive ones. No increase in response to the positive combination was found in the hippocampus. Also, the correlation of activity between the amygdala and the hippocampus was not significantly increased during the positive combination compared with the neutral one. Finally, no clusters of activation were found in limbic regions for the positive conditions in the whole-brain analysis. These results are consistent with previous literature that mainly links the amygdala to negative emotions such as fear (Phan et al. 2002). However, various studies also link the amygdala to appetitive emotions such as sexual arousal, which are usually considered positive (Beauregard et al. 2001; Hamann et al. 2004). It thus seems that the amygdala does not specifically process negative or positive emotions, but rather, as suggested above, emotions related to an immediate behavioral response, such as approach or withdrawal. In the present experiment, scary and joyful music clips were used to induce negative and positive emotions, respectively. Clearly, fear is more directly related to an immediate behavioral response than joy, which may account for the more consistent response that the limbic system shows to films coupled with emotionally negative music in our study.

Several caveats to the finding in the amygdala should be noted. First, it cannot be excluded that the subjects in the fMRI study did experience stronger emotions while viewing the music and film together in comparison with the separate conditions. However, given that in the stimulus judgment experiment subjects reacted similarly, in terms of both valence and arousal, to the combined and separate conditions, we believe this to be unlikely (see Fig. 1). Second, it can also be claimed that the increased amygdalar response to the combination is an effect of salience. The combination of music and film might be more salient than each on its own and thus evoke a greater brain response. What stands against this interpretation is the fact that the combination of film and neutral music did not produce an increased response; rather, an increased response was obtained only when the film was combined with music that attributed an emotional tone to it. Third, it can be noted that the lack of increased amygdalar response to the neutral combination is at least partially a result of an increased response to the neutral film on its own when compared with the films that were combined with negative or positive music. As the different film conditions consist of different films shown to different subjects, the differential response most probably represents individual variability. In fact, all between-valence comparisons in the present study suffer from the same variability, and none of them show statistically significant effects. Therefore, the most reliable way of analyzing the data is by comparing the different conditions only within valence. Using this method, a significant effect was obtained in the amygdala only for the negative and positive combinations and not for the neutral combination. Fourth, although the music clips were generally not familiar to the subjects (as most of them were composed in-house), we must consider that familiarity could have made the combination appear less novel and novelty could have affected the amygdalar response.

**Integrative Emotional Processing in the outside limbic regions: Flexibility and Specificity**

Similar to the amygdala and hippocampus, lateral prefrontal and lateral temporal regions were implicated in the integration of emotions and concrete content (see Table 1). What, then, is their additional contribution? The present study cannot distinguish between the different possible roles that these regions may have in emotional processing. Nevertheless, based on these results and previous research, a sensible model can be proposed. One property that the amygdala by itself seems to lack is flexibility in stimulus–response associations. This was first proposed by Morgan et al. (1993) who showed that extinction of fear-conditioned stimulus–response associations, whose
feeling the real world

role is thought to be in enabling responses do not usually engage the lateral PFC, and its primary formation depends on the amygdala, became difficult after increased response to the negative and neutral music when they have been diluted. Another property that the amygdala lacks is specificity to a particular stimulus type. As suggested by Ledoux, it seems that processing in the amygdala level is done in a relatively crude manner. For example, whereas fear conditioning of an auditory stimulus depends on an intact amygdala, the specificity of the fear response to a distinct tone (and not to other similar sounds) might depend on intact auditory cortex (Jarrell et al. 1987; LeDoux 2000). Also, human amygdalar and hippocampal neurons do not discriminate between particular cases of a general category, such as different houses or different faces (Kreiman et al. 2000). In contrast, lateral temporal regions have been suggested to facilitate processing of particular object concepts, following a posterior–anterior axis of increasingly specific representations, from the broader category level (i.e., faces in general) to the particular object (i.e., Michael Jordan’s face) (Martin and Chao 2001). Thus, it is possible that lateral temporal regions, such as the superior temporal sulcus and the temporal pole, mediate the tying of emotion to representations of particular objects, people, or situations.

**Modality Selectivity in Integrative Emotional Processing**

Our cross-modal approach provided a unique platform for studying modality selectivity in emotional processing. The visual cortex did not show increased activation in response to the combination of auditory emotional stimuli and visual scenes. Furthermore, statistical parametric maps contrasting the emotional combinations with films alone revealed no emotional modulation effect in visual regions. This is in contrast to previous studies that have shown modulation of high-order visual regions by emotion (Rotstein et al. 2001; Phan et al. 2002; Vuilleumier et al. 2004). However, the present study differs from these previous studies in that the emotional character of the visual stimuli in the combination was conveyed through a different modality, suggesting that emotion modulates sensory processing only within the same modality through which it is conveyed. Still, this finding seems to conflict with those of Dolan et al. (2001), which show that emotionally congruent compared with emotionally incongruent visual and auditory stimuli yield enhanced fusiform response. However, this enhanced response may reflect an inhibition that results from incongruence and not enhancement due to congruency, similar to the inhibitory effect that asynchronous audiovisual speech has in visual areas (Calvert et al. 2001). A trend toward decreased activation of sensory regions during multimodal conditions can also be observed in our data (Fig. 7). It is also possible that emotional modulation occurs only in category-specific visual regions during processing of stimuli from that category (e.g., faces in the FFA); as our visual stimuli were not specific to one category, an emotional modulation effect could have been diluted.

Interestingly, the auditory cortex showed a significantly decreased response to the negative and neutral music when they were combined with film clips. This is in line with the findings of Bushara et al. (2002) which showed that cross-modal binding of sound and sight involves decreased activation of unimodal auditory and visual areas coupled with increased activation of multimodal areas. On the other hand, Armony and Dolan (2000) showed that emotional visual context can increase auditory cortex response. However, in their study, context was not manipulated independently of emotion but rather used to elicit it. Therefore, it is possible that it was the resulting emotionality, and not visual context per se, that modulated auditory cortex activity. Further research is needed to clarify the ways through which integration of multimodal information conveying emotionality and context involves unimodal sensory areas.

**Cross-Modal Approach in Affective Neuroscience**

So far, cross-modal methodology has proven successful mostly in locating sites of integration of perceptual information, such as the anterior superior temporal sulcus for audiovisual speech perception (Wright et al. 2003) and the intraparietal sulcus for spatial location (Nobre et al. 1997). We believe that the potential of cross-modal methodology is far greater, in that it offers an effective way for studying conceptual functions. Mesulam (1998) proposed that high cognitive functions invariably involve sites of multimodal convergence. Accordingly, Booth et al. (2002) used a conjunction of visual and auditory stimuli to isolate multimodal regions that are involved in semantic decoding. Following this line of thought, we took the basic, commonly used, cross-modal paradigm and extended it. As done before, stimuli in 2 modalities were presented separately and combined, but instead of having the 2 modalities convey the same conceptual information, here the whole concept emerges only in the combination. An emotion with concrete meaningful content is only evoked when the film and music are combined. The benefits of such a design are substantial. First, a specific conceptual function is isolated from underlying processes as it is only elicited when the 2 modalities are combined. Second, the perceptual features of the stimuli are afforded the ultimate control—identical film and music clips were contrasted in the separate and the combined conditions. Finally, stimuli that are lively, animated, dynamic, and complex like music and films can be used in a strictly controlled experiment addressing a specific hypothesis.

In conclusion, the current study exhibits a possible neural basis for the known contingency between emotion and cognition in humans, keeping in mind that in this study, cognition was confined to perceptual-associated content. Animals and humans share the same basic set of emotions, including emotions such as fear, joy, or sadness (Darwin 1998). However, human cognitive abilities have become extremely sophisticated and involved in all aspects of behaviors including emotions. With such developed cognition, emotions no longer seem like meaningless forces that overrule all other aspects of behavior. For instance, the human phenomenon of anxiety is not merely an emotional expression of fear but rather the maladaptive interpretation associated with the fear in regard to occurrences in the world (Kaplan and Sadock 1998). Freud has proposed that in order to overcome melancholy, one needs to understand its real-world context so that mourning and resolution can take place (Freud 1957). Indeed, in the process of treating emotional problems, psychotherapy often focuses on grounding emotions in the real world of interpersonal relations. The current finding that associated content modulates neural activity even at the
heart of the emotional brain adds support to the fundamental role of real-world content in processing emotions.

Supplementary Material
Supplementary materials can be found at: http://www.cercor.oxfordjournals.org/.

Notes
This study was funded by the Israeli Science Foundation (Bikura Program 1440/04, TH), the Israeli Ministry of Science and Sport. We thank Dr. Pazit Pianka for assistance during data acquisition, Dr. Arron Neufeld and Mr. Oren Levin for technical assistance with data acquisition, Dr. Itamar Kahn for most insightful discussions over the data, Dr. Ricardo Tarrasch for statistical assistance, Dr. David Papo for comments on the manuscript, and Ms. Nirit Binyamini for assistance with the graphics.

This work was done in cooperation with the Levi Edersheim Gitter Institute for Functional Brain Imaging, Tel Aviv University.

This work was performed in partial fulfillment of the MD thesis requirements of the Sakeller Faculty of Medicine, Tel Aviv University. Conflict of Interest None declared.

Address correspondence to Talma Hendler, Tel Aviv Sourasky Medical Center, 6 Weizmann Street, Tel Aviv 61439, Israel. Email: talma@tasmc.health.gov.il.

References


Cerebral Cortex December 2007, V 17 N 12