

ORIGINAL ARTICLE

The Temporal Pole Top-Down Modulates the Ventral Visual Stream During Social Cognition

Corinna Pehrs^{1,2,3}, Jamil Zaki⁴, Lorna H. Schlochtermeyer^{1,2,3}, Arthur M. Jacobs^{1,2,3}, Lars Kuchinke^{1,2,3,5} and Stefan Koelsch⁶

¹Cluster of Excellence “Languages of Emotion”, 14195 Berlin, Germany, ²Department of Education and Psychology, Freie Universität Berlin, 14195 Berlin, Germany, ³Dahlem Institute for Neuroimaging of Emotion, 14195 Berlin, Germany, ⁴Department of Psychology, Stanford University, Stanford, CA 94305, USA, ⁵Department of Psychology, Experimental Psychology and Methods, Ruhr-Universität Bochum, 44801 Bochum, Germany and ⁶Department of Biological and Medical Psychology, University of Bergen, 5009 Bergen, Norway

Address correspondence to Corinna Pehrs, Freie Universität Berlin, Cluster Languages of Emotion, Habelschwerdter Allee 45, 14195 Berlin, Germany. Email: corinna.pehrs@fu-berlin.de; Stefan Koelsch, Department of Biological and Medical Psychology, University of Bergen, 5009 Bergen, Norway. Email: koelsch@cbs.mpg.de

Abstract

The temporal pole (TP) has been associated with diverse functions of social cognition and emotion processing. Although the underlying mechanism remains elusive, one possibility is that TP acts as domain-general hub integrating socioemotional information. To test this, 26 participants were presented with 60 empathy-evoking film clips during fMRI scanning. The film clips were preceded by a linguistic sad or neutral context and half of the clips were accompanied by sad music. In line with its hypothesized role, TP was involved in the processing of sad context and furthermore tracked participants' empathic concern. To examine the neuromodulatory impact of TP, we applied nonlinear dynamic causal modeling to a multisensory integration network from previous work consisting of superior temporal gyrus (STG), fusiform gyrus (FG), and amygdala, which was extended by an additional node in the TP. Bayesian model comparison revealed a gating of STG and TP on fusiform-amygdala coupling and an increase of TP to FG connectivity during the integration of contextual information. Moreover, these backward projections were strengthened by emotional music. The findings indicate that during social cognition, TP integrates information from different modalities and top-down modulates lower-level perceptual areas in the ventral visual stream as a function of integration demands.

Key words: dynamic causal modeling, emotion, empathy, fMRI, multisensory integration

Introduction

The temporal pole (TP) is part of the association cortex and involved in multimodal sensory integration (Olson et al. 2007; Skipper et al. 2011). It has been implicated in various higher-order functions of socioemotional cognition including language processing (Hickok and Poeppel 2007; Altmann et al. 2012), face processing (Jimura et al. 2009), emotion (Royet et al. 2000; Aust et al. 2013), and empathic behavior (Rankin et al. 2006; Parkinson

and Wheatley 2014). The exact function of TP, however, has remained equivocal.

One prominent theory about TP functioning is the “semantic hub” theory, which considers TP as a domain-general hub integrating semantic information from different modalities into a coherent representation (McClelland and Rogers 2003; Patterson et al. 2007). This account is based on findings from patients with semantic dementia, showing that neurodegenerative loss in the anterior temporal lobe (aTL) is associated with

progressive deficits in recognizing objects across modalities (Binney et al. 2010; Visser et al. 2010; Lambon Ralph et al. 2012).

The TP has extensive connections to other brain regions, which foster its putative neuromodulatory impact consistent with the semantic hub account (Patterson et al. 2007) and with the characteristics of a convergence region (Damasio 1989). Structurally, it is located at the end of the ventral visual stream and strongly interconnected with the amygdala (AMY; Nakamura and Kubota 1996; Stefanacci and Amaral 2002) as well as with antecedent regions of the ventral visual stream via the longitudinal fasciculus (Crosby 1963; Catani et al. 2003).

The ventral visual stream is involved in object recognition and assigns conceptual knowledge and meaning to visual and auditory information (Carlson et al. 2014). It is characterized by a posterior–anterior gradient of increasing receptive field size and increasing complexity of the representations posing TP as a likely region to exert top-down modulations via backward projections to lower-level regions of the stream (Gilbert and Li 2013; Kravitz et al. 2013). Crucially, there is evidence that these backward projections are modulated by integration demands during visual and linguistic object recognition (Bar et al. 2006; Chan et al. 2011; Campo et al. 2013; Yvert et al. 2012). Thus, top-down projections from more complex to lower-level perceptual areas help to optimize recognition of sensory input (Bar 2003; Friston 2010). This phenomenon has been investigated within the framework of facilitation models (Ullman 1995; Bar 2003; Bar et al. 2006) describing how object recognition is improved by integration of higher-order representations with lower-level sensory processing via top-down modulations.

To adopt these findings to social cognition and to test whether TP acts as a semantic hub integrating over complex social cues, we employed naturalistic stimuli: Empathy-evoking movie sequences depicting protagonists undergoing emotional experiences. Sharing such experiences with filmed protagonists requires a continuous neural multisensory integration of visual, auditory, and contextual information (Raz et al. 2014).

This integration was investigated here using dynamic causal modeling (DCM), which in conjunction with naturalistic stimuli provides a highly suitable method and an elegant test of the hub account and top-down facilitation in the social domain, where powerful connectivity analyses are still rare. Effects of multisensory combination (visual, auditory, and contextual) on changes of effective connectivity were tested in a multisensory integration network. The network structure was built upon our previous work on the modulatory impact of emotional music on the perception of positively valenced kissing scenes from romantic comedies (Pehrs et al. 2014) and by the results of the present study. In the previous study (Pehrs et al. 2014), the anterior superior temporal gyrus (aSTG) was identified as a convergence region integrating multisensory (visual and auditory) emotional information within auditory cortices. Using nonlinear DCM, it was demonstrated that the aSTG controls visual–limbic connectivity from fusiform gyrus (FG) to AMY, both of which are involved in congruent multisensory stimulation (Dolan et al. 2001; Baumgartner et al. 2006; Eldar et al. 2007). These regions (aSTG, FG, and AMY) were again responsive to multisensory compared with unisensory stimulation in the present study.

Both the TP and the aSTG are suggested as convergence regions with reciprocal projections to the sensory association areas (Damasio 1989). To examine the influence of TP as a second convergence region, which is particularly involved in the processing of semantic context information, we used sad empathy-evoking film clips and context variations as an additional experimental manipulation besides music. The network

architecture from the previous study was extended in the present study by the TP resulting in a four-region integration network (aSTG, FG, and AMY + TP). The aim was to (1) replicate a gating effect of the aSTG on fusiform–amygdalar connectivity during the perception of negatively valenced film stimuli, and to (2) test a putative gating of TP with neuromodulatory influence on the same fusiform–amygdalar connectivity. Moreover, we were particularly interested (3) in how top-down semantic information changes the connectivity from TP to AMY and FG. Consistent with top-down facilitation models and findings from object recognition, we hypothesized backward projections from TP to prior stages of the ventral visual stream (FG) in dependence of integration demands varied by accumulating information from different modalities.

Materials and Methods

Participants

Twenty-eight healthy subjects participated in the study. Owing to head movements over voxel size and 3°, 2 participants had to be excluded leaving 26 subjects (17 females and 9 males, mean age 30.3 ± 8.7 years) for the final analysis. All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield 1971) and had normal or corrected-to-normal vision. After a general screening for MR compatibility including exclusion of any neurological or psychiatric disorders, participants were informed about the study and written consent was obtained. The study was approved by the ethics committee of the German Psychological Society. Participants either received course credit or were paid for their participation.

Visual, Auditory, and Context Stimuli

The fMRI design includes film clips (“visual”), pieces of music (“auditory”), and short text passages (“context”), which are described in the following section. The selection of visual stimuli was based on film theory. Plantinga (1999) describes a specific sort of scene, “the scene of empathy,” which is applied intentionally by film directors to elicit empathetic emotions in the viewer. This sort of scene utilizes a prolonged shot of the character’s face to focus the attention on his or her interior emotional experience. Based on Plantinga’s descriptions, 60 film stimuli were selected from feature films, which depict a close-up of a character (30 males and 30 females) for 12 s with a sad facial expression and no mouth movements (for a list of film stimuli, see [Supplementary Table 1](#)). By using sad characters, the stimuli were intended to evoke empathic concern. They were also used in another study investigating the role of attention during up- and down-regulation of empathic concern (Manera et al. 2014). A large set of film clips ($n = 113$) were rated in advance with regard to their emotional quality [i. e., arousal, happiness, sadness, compassion, familiarity, liking of characters on a scale from 1 (not at all) to 7 (very much), and valence on a scale from -3 (very unpleasant) to 3 (very pleasant)] by an independent sample of 13 participants (9 females, age 29.3 ± 5.9). The purpose was to select a congruent, homogeneous set of equally emotion-evoking stimuli. The film clips of the final selection were perceived as moderate on emotion scales (valence -0.36 ± 0.49 , arousal 3.55 ± 0.51 , sadness 4.06 ± 0.86 , and compassion 3.80 ± 0.49), which ensured the ability for modulation by music and context. Post-scanning ratings on a scale from 1 (known) to 4 (unknown) confirmed that the film clips were not familiar to the participants (3.74 ± 0.19). Some film clips were minimally close-cropped using the film editing Software Adobe Premiere

Pro CS5.5 (Adobe System Incorporated®, San Jose, CA, USA) to only show the face and the shoulders of the person. All film clips were faded in and out with a black shade of 0.5 s and saved with a resolution of 600 × 800 pixels.

The “auditory stimuli” consisted of 40 sad music pieces [similar to the ones used in Pehrs et al. (2014)]. Sad music was used to increase the emotional intensity of the sad film clips. All auditory stimuli (except for one) were in minor mode and instrumental (for a list of auditory stimuli, see Supplementary Table 2). The waveforms were root mean square-normalized using the digital audio software Adobe Audition CS5.5 (Adobe System Incorporated®) and the fading properties were matched according to the video shades.

The “context stimuli” consisted of short written texts and were used to place the character in a narrative situation before each film clip was presented (henceforth, referred to as context). For every film clip, a sad and a neutral context were created. The sad context was based on the actual situation of the character in the movie and described difficult psychological circumstances (e.g., the recent loss of a loved person and getting informed about one’s infertility). The neutral context was created describing the general set-up, in which the character is presented in the clip (e.g., sitting in a car and drinking a glass of water), to control

for processing of nonemotional language. All contexts consisted of 2 sentences (20.0 ± 2.1 words, range 17–23) and were matched with regard to sentence structure, word number, and reading time (see Fig. 1 for examples).

A large set of auditory ($n = 60$) and context stimuli ($n = 226$) was rated in advance by another independent sample of 26 participants (19 females, age 28.38 ± 7.27) on various emotion scales [valence, arousal, happiness, sadness, compassion, imageability, and congruity of visual, auditory, and context stimuli from 1 to 7 (valence -3 to 3)]. This was done to select a most emotionally congruent set of stimuli (sadness of sad contexts 5.25 ± 0.68 and sadness of music 4.24 ± 0.47) with highest significant differences to the neutral contexts (P -values for all emotion dimensions < 0.001 , Bonferroni-corrected). All selected stimuli were perceived as congruent to the film clips (sad contexts 4.60 ± 1.21 ; neutral contexts 4.04 ± 0.8 ; and music 5.08 ± 0.81).

Experimental Design

To investigate different neural activity induced by multisensory integration and contextual framing, the multisensory combination of the same visual information (empathic film clips) was systematically varied by music and context in a 2×3 design.

A Experimental design

		Context		
		No	Neutral	Sad
Music	No	1	2	3
	Sad	4	5	6

B Conditions

- ① Film only
- ② Neutral context, film + no music
- ③ Sad context, film + no music
- ④ No context, film + music
- ⑤ Neutral context, film + music
- ⑥ Sad context, film + music
- ⑦ Music only

C Examples of conditions

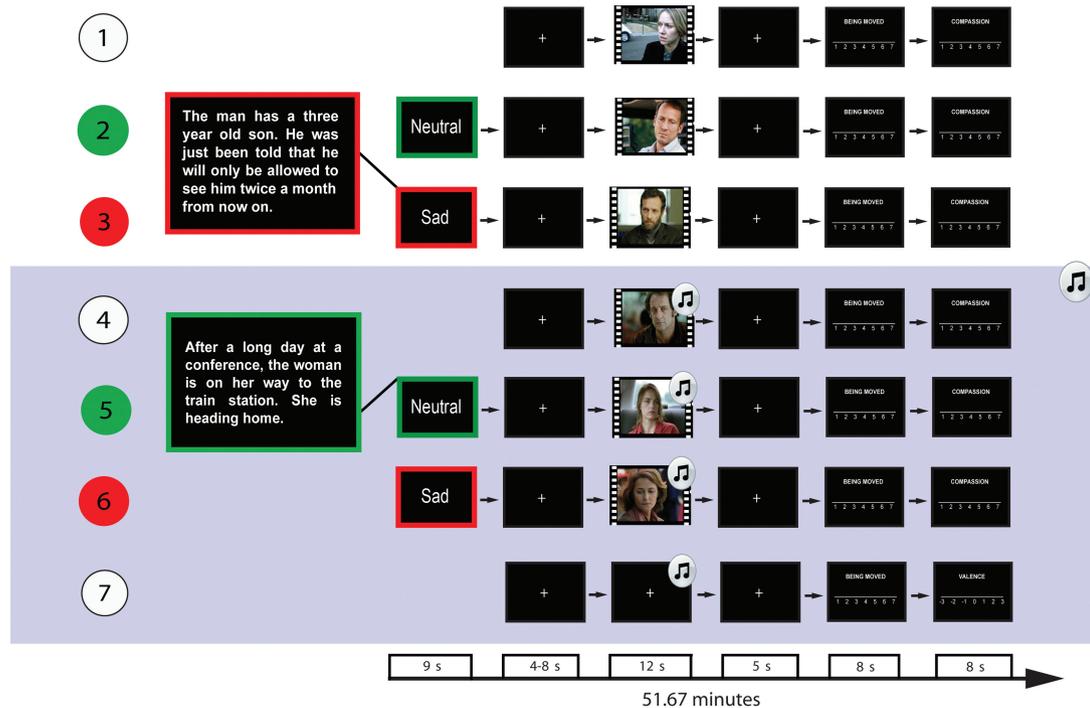


Figure 1. Factorial design and example of stimuli. (A) A 2×3 experimental design schematically depicting the 2 factors music (rows, no/sad) and context (columns, no/neutral/sad). (B) Legend of conditions with no music (left) and with sad music (right). (C) Examples illustrating the presentation order within each trial type. A fully randomized design was employed to present 70 trials in total (10 trials per condition).

The resulting 6 conditions are illustrated in Figure 1. Each condition included a 12-s empathic film clip preceded by a fixation cross with varying duration (4–8 s, mean 6.19 s). The first condition (Fig. 1C1) showed a silent film clip without preceding context (“no context” condition = film only). The second condition (Fig. 1C2) started with a neutral context followed by a silent film clip (“neutral context” condition = neutral context, film + no music). The third condition (Fig. 1C3) started with a sad context followed by a silent film clip (“sad context” condition = sad context, film + no music). Conditions 4, 5, and 6 also arose from the 3 different context conditions (no/neutral/sad context), but included sad music played during the presentation of the film clip. Thus, the 6 conditions shown in Figure 1C are the result of all possible combinations of the factor context (no/neutral/sad) with the factor music (no/sad; Fig. 1A). To be able to determine neural activations due to multi- versus unisensory processing, a purely auditory control condition was included, in which sad music was presented during a 12-s fixation cross without the influence of meaningful visual or context information (Fig. 1C7), resulting in 7 conditions in total. After every trial in each condition, a 5-s fixation cross was presented followed by 2 ratings, in which the participants used a tracking ball to rate their current emotional state in terms of “compassion” (“valence” for music only) and “being moved” on a 7-point Likert scale (1 not at all, 7 very much) for 8 s each. For the purpose of this study and for the sake of brevity, we focused on the participants’ experience of compassion as a core measure for socioaffective processing. Compassion is a prosocial oriented subspect of empathy (Zaki and Ochsner 2012) and is often defined through both affective (feeling for someone suffering) and motivational (desire to help) components (Goetz et al. 2010). In the instruction for the participants, we focused on the affective component. Ten trials per condition resulted in 70 trials presented in a fully randomized order during 51.67 min of continuous fMRI data acquisition. The 40 music pieces were presented randomly with the film clips to avoid multiple presentations of one film clip with a specific piece of music. All stimuli were presented using the stimulation software Presentation (Version 9.00, Neurobehavioral Systems, Albany, CA, USA). The visual stimuli were presented using LCD goggles (Resonance Technology, Northridge, CA, USA). Auditory stimuli were presented binaurally through MRI-compatible headphones with a standard comfortable listening level of 70 dB. Before the fMRI measurement, participants were instructed to pay attention to the film clips and to attribute the context to the character. They were trained with 5 test trials inside the scanner.

Functional Magnetic Resonance Imaging

The imaging data were acquired with a 3-T Siemens (Erlangen, Germany) Tim Trio MRI scanner at the D.I.N.E. (Dahlem Institute for Neuroimaging of Emotion, Freie Universität Berlin, Germany) using a 12-channel phased-array headcoil. After a high-resolution T_1 -weighted structural image for registration of the functional data [time repetition (TR) 1900 ms; time echo (TE) 2.52 ms; flip angle 9° ; voxel size 1 mm^3 ; 176 sagittal slices; and 1 mm slice thickness], a whole brain T_2^* -sensitive gradient echo-planar imaging sequence was applied (TR 2000 ms; TE 30 ms; voxel size 3 mm^3 ; 1356 scans; flip angle 70° ; field of view 192 mm^2 ; matrix 64×64 ; 37 slices; 3 mm slice thickness; and 0.6 mm gap).

Analysis of Behavioral Data

Behavioral data were analyzed using SPSS20 (SPSS, Inc., Chicago, IL, USA). A 2×3 repeated-measures ANOVA was performed with

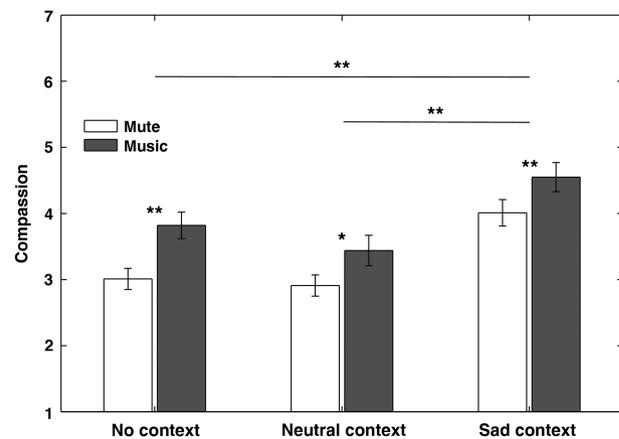


Figure 2. Results of behavioral data. The emotional properties of the clips were rated in terms of compassion using a 7-point Likert scale (1 not at all, 7 very much). Music and sad context significantly increased the ratings. Error bars represent standard error of the mean. * $P < 0.05$, ** $P < 0.001$.

the factor music (no/sad) and context (no/sad/neutral) on ratings of compassion. Post hoc paired t-tests were performed and results were considered statistically significant when $P < 0.05$ after Bonferroni correction for multiple comparisons. The results are presented in Figure 2.

General Linear Model Analysis

fMRI data were analyzed using the Statistical Parametric Mapping software package SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK; <http://www.fil.ion.ucl.ac.uk>, last accessed September 27, 2015) implemented in MATLAB (version 2011a; The MathWorks, Inc., Natick, MA, USA). Before preprocessing, the origin of the functional time series was set to the anterior commissure. Motion correction was performed using each subject’s mean image as a reference for realignment. The T_1 image was coregistered to the mean functional image generated during realignment. The coregistered T_1 images were segmented with the “New Segment” routine and normalized using DARTEL tools (Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra). The images were resliced with a 3-mm isotropic voxel size and smoothed with a 6-mm full-width at half maximum isotropic Gaussian kernel.

Statistical analysis was performed using the general linear model with a two-level approach (Friston et al. 1995). On the first level, a design matrix was fitted for each subject with one regressor for each condition (film only; neutral context, film + no music; sad context, film + no music; no context, film + music; neutral context, film + music; sad context, film + music; music only; Fig. 1B). To control for effects related to reading and comprehension, the presentation of sad and neutral contexts and the fixation crosses appearing after the contexts were additionally included as 4 regressors. Taken together, they model the time before film clip presentation. To account for movement-related variance, the realignment parameters were added as regressors in the design matrix. All events were modeled as a boxcar with the duration of the event and convolved with the standard hemodynamic response function.

To localize brain activations proving the efficiency of our experimental factors, context and music, as well as their interactions, the 6 contrast images corresponding to conditions 1–6 (Fig. 1B) versus baseline were taken on the second level to a

mixed 2 × 3 factorial ANOVA for random-effects (RFX) analysis. To identify neural activations of an emotion effect for sad contexts, activity based on sad contextual framing was contrasted against activity based on neutral contextual framing ([sad context, film + music&sad context, film + no music] > [neutral context, film + music&neutral context, film + no music]; Table 1). To localize brain regions engaged in multisensory integration, a contrast was calculated comparing bimodal versus unimodal presentations ([no context, film + music] > [film only&music only]; Table 2).

A second design matrix was constructed for 24 participants (2 participants were excluded due to missing values in behavioral data) to perform a parametric modulation of the blood oxygen level-dependent (BOLD) signal as a function of experienced emotion. This matrix consisted of the 6 conditions: film only; neutral context, film + no music; sad context, film + no music; no context, film + music; neutral context, film + music; sad context, film + music (Fig. 1B1–6). For every condition, we included a second column in the model (parametric modulators) containing the subjects' individual ratings of compassion for each of the 10 video clips in that condition. This analysis was implemented to test

for a linear relationship between regional signal changes depending on subjectively experienced empathic concern (compassion ratings) for the character. We report BOLD activity responsive to an increase in empathic concern, combined over all conditions. By default, SPM applies a serial orthogonalization to parametric modulators. This model is henceforth referred to as “design matrix with parametric modulation” whereas the model described before is referred to as “design matrix without parametric modulation”. All contrast images were taken to the second level for an RFX analysis. We first identified all clusters yielded in the uncorrected SPMs with a threshold of $P < 0.001$ [and a cluster-extent of 30 voxels], and then—to exclude all false-positive clusters—identified and reported only those clusters which were significant at the cluster level with FWE correction ($P < 0.05$).

Based on results showing the involvement of TP in socioemotional cognition (Zahn et al. 2007, 2009; Simmons et al. 2010; Willems et al. 2011; Olson et al. 2013), activation in the TP was hypothesized a priori. Moreover, TP activation is likely missed because of the susceptibility artifacts in the BOLD signal in the region of the aTL (specifically the ventral part) due to the

Table 1 Cortical activations of sad versus neutral context and parametric modulation with compassion ratings

Region	Brodmann area	MNI coordinates			Cluster size	T-value
		x	y	z		
[Sad context, film + music&sad context, film + no music] > [neutral context, film + music&neutral context, film + no music] ^a						
R. temporal pole	BA38	39	12	−36	10	3.16
L. temporal pole	BA38	−42	9	−33	10	3.49
Parametric modulation with compassion ratings						
R. temporal pole	BA38	51	12	−27	62	5.04
R. temporal pole	BA38	54	6	−12	—	4.21
R. superior temporal gyrus	BA22	48	−33	3	54	4.86
R. superior temporal gyrus	BA22	48	−21	−3	—	4.41
R. middle temporal gyrus	BA22	57	−24	0	—	3.52

Note: Clusters were obtained using a voxel threshold of $P < 0.001$, a cluster-extent of 30 voxels and an FWE-corrected cluster threshold of $P < 0.05$. Additional local maxima that are at least 8 mm distant from the main peak are listed.

^a $P < 0.05$, small-volume FWE-corrected.

Table 2 Regional brain activation during multi- compared with unisensory stimulation

Region	Brodmann area	MNI coordinates			Cluster size	T-value
		x	y	z		
[No context, film + music] > [film only&music only]						
R. cuneus	BA17	6	−81	6	11 857	18.73
L. cuneus	BA18	−3	−93	12	—	16.70
L. lingual gyrus	BA18	−3	−81	−6	—	16.04
R. middle frontal gyrus	BA6	45	9	45	246	6.65
R. precentral gyrus	BA9	36	9	39	—	6.47
R. middle frontal gyrus	BA9	36	9	27	—	5.80
R. precuneus	BA7	24	−51	54	48	6.46
R. inferior frontal gyrus	BA47	54	39	−6	102	6.10
R. inferior frontal gyrus	BA47	45	33	−6	—	4.60
R. inferior frontal gyrus	BA46	57	36	12	—	4.51
L. medial frontal gyrus	BA11	−3	57	−12	52	5.98
R. medial frontal gyrus	BA11	6	57	−12	—	4.87
R. medial frontal gyrus	BA11	3	48	−15	—	4.84
R. postcentral gyrus	BA3	21	−36	57	191	5.96
R. postcentral gyrus	BA4	30	−33	60	—	5.82
R. postcentral gyrus	BA3	21	−36	66	—	5.57
L. parahippocampal gyrus		−21	−9	−21	42	4.95

Note: Clusters were obtained using a voxel threshold of $P < 0.001$, a cluster-extent of 30 voxels and an FWE-corrected cluster threshold of $P < 0.05$. Additional local maxima that are at least 8 mm distant from the main peak are listed.

proximity of air-filled sinuses and ear canals (Devlin et al. 2000; Binney et al. 2010; Visser et al. 2010). Accordingly, small-volume corrections ($P < 0.05$, FWE-corrected at the voxel level) were applied using bilateral volumes of interest of the TP taken from the WFU Pick Atlas (Lancaster et al. 2000; Maldjian et al. 2003).

The results are reported in Tables 1 and 2, and Figure 3.

Dynamic Causal Modeling

Overview

To examine context- and music-dependent changes in network dynamics during multisensory integration, we performed an effective connectivity analysis using DCM10 (Friston et al. 2003) as implemented in SPM8 (update r5236). By adjusting the parameters of a predicted neuronal system and measured BOLD signal (time series), DCM allows inferences on dynamics of interacting brain regions on a neural (hidden) level and specifically on perturbations by experimental conditions [“modulatory input” (1)]; here music and context. Furthermore, it distinguishes between endogenous coupling [“intrinsic connections” (2)] and direct driving of regional activity [“driving input” (3)]. By including a

nonlinear term to model the influence of activity in one region on the coupling of 2 other brain regions (Stephan et al. 2008), we examined a gating effect on fusiform–amygdalar connectivity by aSTG (Müller et al. 2012; Pehrs et al. 2014) and by TP. Therewith, we tested whether neuronal populations that converge information over modalities (aSTG and TP) control the coupling of 2 other brain regions (FG and AMY) engaged in emotion processing of multisensory stimulation (Dolan et al. 2001; Pehrs et al. 2014). Neurobiologically, gating effects represent indirect physiological modulations mediated by synaptic connections from one or more regions (aSTG and/or TP) on the connectivity between 2 other regions (FG and AMY).

In DCM for fMRI, the modeled neuronal dynamics are transformed to a hemodynamic response by using a hemodynamic forward model (Friston et al. 2000; Stephan et al. 2007). Parameter estimation is performed in a Bayesian framework as described previously (Friston et al. 2003). For selection of the best-fitting model, Bayesian model selection (BMS) was performed. BMS takes into account the fit of the models, primarily based on the number of free parameters, in relation to the model complexity. With higher complexity, the relative fit of a model may increase,

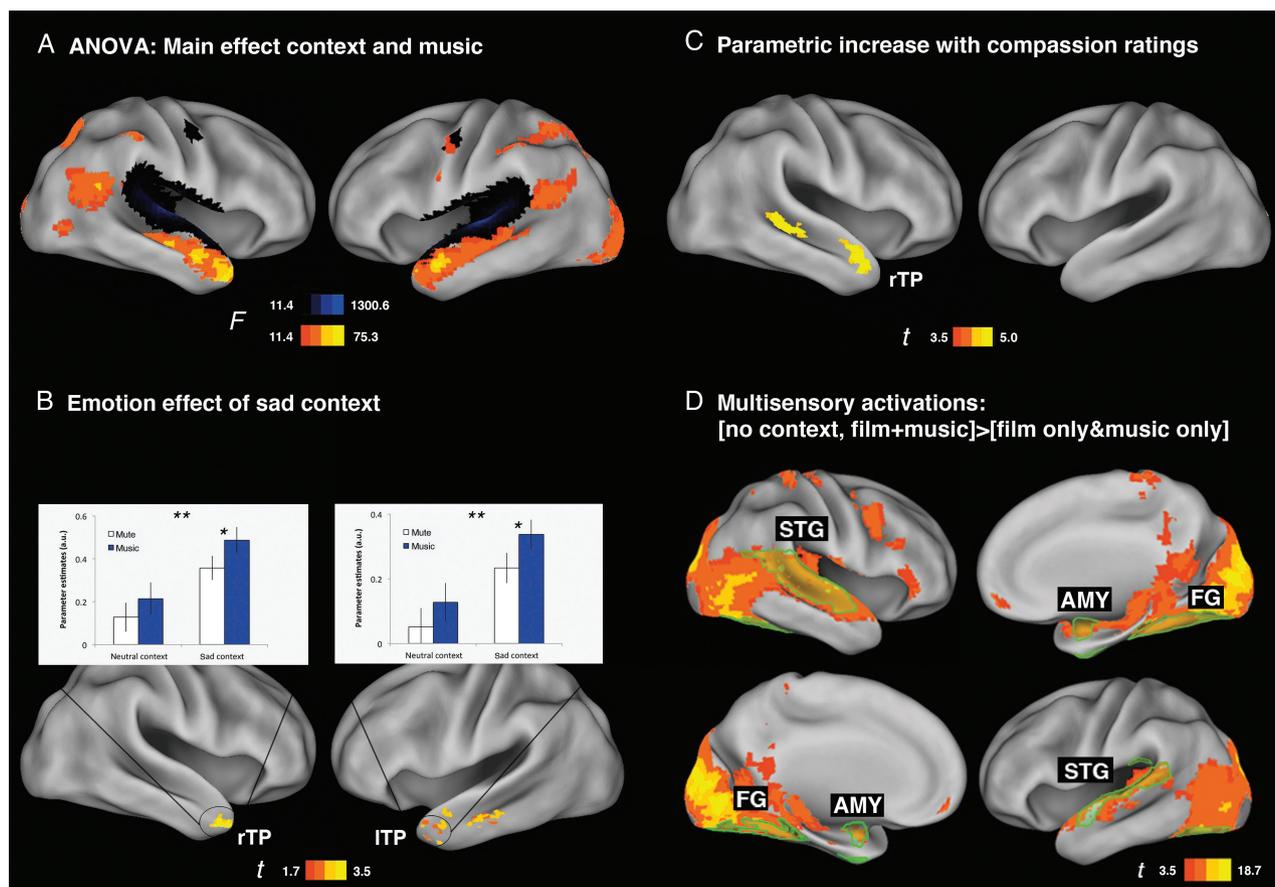


Figure 3. Regional brain activations during social cognition. (A) Music by context ANOVA: Main effect of context (red) and main effect of music (blue). (B) Emotion effect of sadness during presentation of the film clips [sad context, film + music&sad context, film + no music] > [neutral context, film + music&neutral context, film + no music]. For illustration, activity is thresholded at $P < 0.05$ (uncorrected) and a cluster extent of 10 voxels. Bar plots represent parameter estimates in arbitrary units (a.u.) within a 6-mm sphere around individual peaks (error bars: SEM). * $P < 0.05$, ** $P < 0.001$. (C) Parametric modulation by compassion ratings show activations in STG/middle temporal gyrus and anterior temporal lobe in the right hemisphere. (D) Multisensory compared to unisensory stimulation [no context, film + music] > [film only&music only] on lateral and medial views of the right hemisphere (upper panel) and left hemisphere (lower panel) overlaid with anatomical ROIs taken from WFU Pickatlas (green) showing activations in brain regions of interest for DCM analysis, the superior temporal gyrus (STG), the amygdala (AMY) and the fusiform gyrus (FG). Activations of A, C and D depicted at a voxel threshold of $P < 0.001$, a cluster-extent of 30 voxels and an FWE-corrected cluster threshold of $p < 0.05$. The maps are projected onto a 3D brain surface using the Caret5 software.

but generalizability may be reduced. RFX BMS gives so-called exceedance probabilities (EPs), the probability that one model is more likely than another. The BMS version integrated in DCM10 for SPM8 also allows BMS on family levels to identify families of models, which share a specific characteristic and are more likely than others not to have that characteristic (Penny et al. 2010).

Regions and Time Series Extraction

Pehrs et al. (2014) identified the aSTG as a multisensory convergence region for emotional information with a gating effect on FG to AMY connectivity, both of which are involved in emotion processing during audiovisual stimulation (Dolan et al. 2001; Jeong et al. 2011). This multisensory integration network of emotion (aSTG, AMY, and FG) was extended in the present study by a socioemotional component in the TP. First, a mixed 2×3 (music \times context) factorial ANOVA (based on the design matrix without parametric modulation) revealed a main effect of context in TP (Supplementary Table 3 and Fig. 3A). Second, an emotion effect for sad versus neutral context was associated with activations in the TP as shown by the T-contrast [sad context, film + music & sad context, film + no music] > [neutral context, film + music & neutral context, film + no music] (based on the design matrix without parametric modulation; Table 1 and Fig. 3B). Third, activation in the TP shows a parametric increase in dependence of experienced compassion over all conditions, which is revealed by design matrix with parametric modulation (Table 1 and Fig. 3C). The other regions (aSTG, AMY, and FG) of the network were engaged for multi- compared with unimodal presentation (Table 2 and Fig. 3D). Taken together, all 4 nodes of the DCM model were responsive to the experimental manipulations of the task (for mean fMRI response across all conditions, see Supplementary Fig. 1).

Regional time series were extracted on the single-subject level using the design matrix without parametric modulation. Each subject's MNI coordinates of the highest T-value within regions of interest (ROIs) of the dynamic causal models (aSTG, FG, AMY, and TP) were surrounded with a sphere of 6 mm and used to extract the first eigenvariate. The ROIs were carefully constructed using a combination of functional and anatomical criteria. Second-level analysis cluster revealed by the T-contrast for multisensory integration ([no context, film + music] > [film only & music only]; for FG, AMY, and aSTG) was masked with corresponding anatomical ROIs (FG, AMY, and STG) taken from the WFU Pick Atlas toolbox (Lancaster et al. 2000; Maldjian et al. 2003). Consistently, the activation cluster in the TP, revealed by the main effect of context in the mixed 2×3 (music \times context) factorial ANOVA, was masked with an anatomical region of the TP taken from the WFU Pick Atlas.

Model Space

The four-area DCM was specified for all subjects with endogenous connections between the regions (aSTG, FG, AMY, and TP) in both directions (Fig. 4). The conditions containing a film clip were defined as driving input to FG (Fig. 1B1–6), the visual input region of the model. The conditions containing sad music (Fig. 1B4–7) were defined as driving input entering the STG, the auditory input region of our model. To test whether context information directly drives TP, all context conditions were additionally modeled as input for direct synaptic responses in this region. While in Figure 4 the modulatory input is subsumed under the single term “context,” it comprises all 4 context conditions (sad context, film + music; neutral context, film + music; sad context, film + no music; neutral context, film + no music; Fig. 1B2, 3, 5,

and 6), allowing us to test the modulatory effects of these conditions (on connectivity) separately.

The basic model structure including intrinsic connections and driving inputs can be seen in Figure 4A. This model was systematically varied with regard to bilinear and nonlinear terms. In the bilinear model family (Fig. 4B), we asked for the directionality of how context and music affects the connectivity strength from TP to AMY for emotional associations and TP to FG for visual associations with all possible combinations, resulting in 15 models. Assuming that areas integrating multisensory information control connectivity of other brain regions (Damasio 1989), this bilinear model space was modeled with gating effects of aSTG, TP, or both on fusiform-amygdalar connectivity (Fig. 4C1–5).

The first gating family assumed a gating effect of aSTG only (Fig. 4C1). The second and third gating family assumed a gating of aSTG in combination with a gating of TP on either AMY to FG (Fig. 4C2) or on FG to AMY (Fig. 4C3). The following 2 families assumed a gating effect of TP only on AMY to FG (Fig. 4C4) or FG to AMY (Fig. 4C5). In total, 90 models were stratified in 6 model families, of which 5 contained a nonlinear term. BMS was performed for the right and in the left hemisphere separately.

Bayesian Model Selection

A comparison of the model space was performed using BMS on 3 family levels (Penny et al. 2010). This was done to reduce the model space systematically stepwise following our leading hypothesis. On the first family level, we compared models with versus without nonlinear modulation (Fig. 4B vs. C). In line with previous results, we expected a better fit of models comprising a gating effect. If so, we aimed to explore on a second level whether the gating is performed by aSTG only or also by TP (Fig. 4C1 vs. 2–5) in correspondence with the hypothesized neuro-modulatory role of TP (Damasio 1989; McClelland and Rogers 2003). On a third family level, the remaining 4 families (2 with gating effect of the TP only and 2 with a bi-regional gating of aSTG and TP) were compared (Fig. 4C2–5). We expected to replicate a gating of aSTG and an additional gating of TP. Finally, the models with bilinear modulation within the best-fitting family were compared to test how effective connections of TP are modulated by top-down information presented before sensory cues. At this final level, we particularly expected feedback projections from TP to FG, consistent with findings from object recognition (i.e., top-down facilitation models) and with Miller and Cohen's (2001) theory of feedback projections as core mechanisms of top-down control.

Because the modulatory input was based on cognitive processes, we expected intersubject variability and used the RFX approach for group analysis (Stephan et al. 2009). For inference on connectivity parameters, standard RFX analysis (ANOVA, one-sample t-tests) was performed.

Results

Behavioral Results

An ANOVA with “compassion ratings” revealed a significant main effect of music ($F_{1,24} = 72.75, P < 0.001$), a significant main effect of context ($F_{2,48} = 31.65, P < 0.001$), and no significant music by context interaction ($F_{2,48} = 1.28, P = 0.332$). Paired t-tests showed that music led to higher compassion ratings in all context conditions (no context: $t_{(25)} = 7.46, P < 0.001$; neutral context: $t_{(25)} = 4.36, P = 0.001$; and sad context: $t_{(24)} = 4.80, P < 0.001$; Fig. 2) and that the sad context significantly increased the compassion ratings compared with the neutral context [$t_{(24)} = 6.16, P < 0.001$; Fig. 2] and

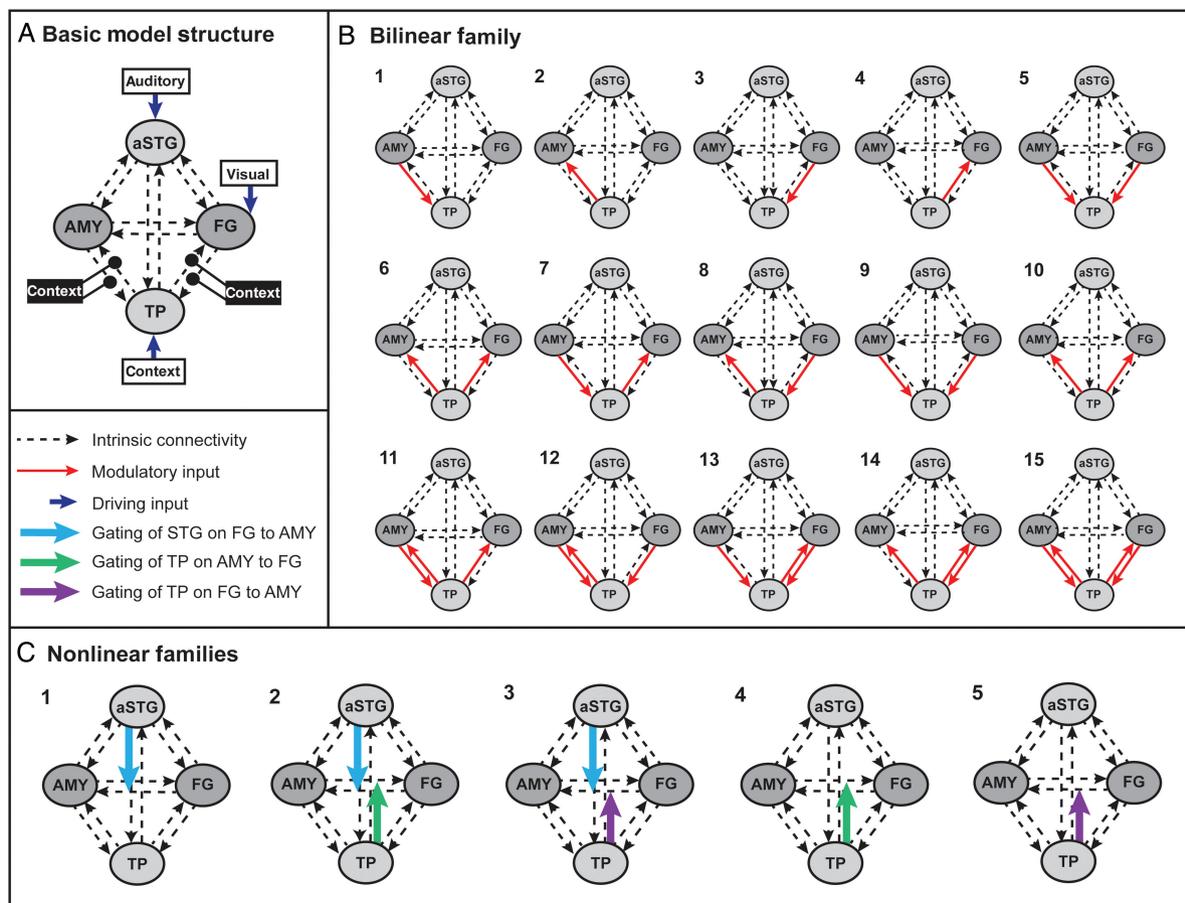


Figure 4. Model space. (A) Basic model structure with nodes, intrinsic connections (dash dotted arrows), and driving input (dark blue arrows). The “visual” driving input on FG includes all conditions containing a film clip (Fig. 1B1–6). The “auditory” driving input includes all conditions containing music (Fig. 1B4–7). The “context” driving input includes all conditions containing context (Fig. 1B2, 3, 5, and 6). For clarity, the modulatory input is labeled “context,” which in fact contains the 4 distinct modulatory inputs (neutral/sad context followed by a silent film clip or a film clip accompanied by music). (B) Bilinear model family with all possible bilinear modulations on TP–FG and TP–AMY coupling in both directions. (C) Nonlinear families; the bilinear model family was tested in all of the nonlinear families resulting in 90 models per subject.

with the no context condition [$t_{(24)} = 5.83, P < 0.001$; Fig. 2]. Taken together, the results indicate that both sad music and sad context significantly increased the experienced compassion of the participants.

fMRI Results

Neural Correlates of Music and Context

We conducted a mixed 2×3 factorial ANOVA (based on the design matrix without parametric modulation) to examine brain regions responsive to our main experimental manipulations, context and music. The results of this ANOVA confirmed the effectiveness of our experimental manipulations: There were significant main effects (FWE-corrected $P < 0.05$ at the cluster level) of music in primary and secondary auditory cortices, the precentral gyrus, cerebellum, and midbrain (Fig. 3A and Supplementary Table 3). Significant main effects of context were found in primary visual areas, temporo-parietal junction (TPJ), middle temporal gyrus, and in the TP (Fig. 3A and Supplementary Table 3), consistent with the ventral visual pathway (Mishkin et al. 1983; Ungerleider and Haxby 1994) and the ventral stream of language processing (Hickok and Poeppel 2004), both implicated in the processing of meaning (Jeannerod and Jacob 2005; Carlson et al. 2014). Further activations of context were found in the precuneus, medial

frontal gyrus, right hippocampus, and cerebellum (Supplementary Table 3). There was no significant music by context interaction.

Neural Correlates of Sad Context

To investigate an emotion effect of context, we examined brain regions engaged in the processing of sad compared with neutral contexts (based on the design matrix without parametric modulation). This was done by contrasting activations during the presentation of the film clips that were preceded by a sad compared with a neutral context ([sad context, film + music&sad context, film + no music] > [neutral context, film + music&neutral context, film + no music]). As hypothesized, brain activations associated with the presentation of a sad context were found in the TP bilaterally (Fig. 3B and Table 1).

Parametric Effect of Compassion Ratings

The involvement of the TP in emotion processing of social stimuli is crucially supported by the parametric analysis (based on the design matrix with parametric modulation). This revealed a linear change of BOLD signal in the right TP and right superior and middle temporal gyrus as a function of experienced empathic concern (Fig. 3C and Table 1).

Neural Correlates of Multisensory Integration

Results from previous studies suggested a stronger engagement of the STG, AMY, and FG during multisensory integration (Dolan et al. 2001; Baumgartner et al. 2006; Eldar et al. 2007; Kreifelts et al. 2007). These regions were the nodes of the network examined in our previous study (Pehrs et al. 2014) and again modeled in the network of the present study. Their involvement in multisensory integration was further supported by the univariate results (based on the design matrix without parametric modulation) and used to build anatomical ROIs in order to extract time courses for the DCM analysis. The contrast [no context, film + music] > [film only & music only] (based on the design matrix without parametric modulation) revealed large clusters of neural activity covering primary and secondary visual and auditory cortices as well as the aSTG, the FG, and the AMY (FWE-corrected $P < 0.05$ at the cluster level; Fig. 3D and Table 2). Consistent with a study by Eldar et al. (2007), activations in the hippocampus and right precentral gyrus were further responsive to a multisensory combination. Additional activations were found in the precuneus, cerebellum, superior and middle temporal gyrus, ventral medial prefrontal cortex, and right inferior frontal gyrus (Fig. 3D and Table 2).

DCM Results

RFX BMS on family level gives so-called EPs, that is, the probability of one class of models is more likely than another class or classes of models in the comparison. On the first family level for the right hemisphere, the nonlinear families clearly outperformed the bilinear family (family EP 99.95%; Fig. 5A), showing that the fusiform–amygdalar connectivity is gated by convergence regions (aSTG and/or TP). On the second family level, gating families including TP outperformed the gating family with aSTG as sole gating structure (EP 74.31%; Fig. 5A) showing the involvement of TP in controlling fusiform–amygdalar coupling. BMS on the third family level determined the winning model family, that is the family with a gating of aSTG and TP on fusiform to amygdalar connectivity (EP 46.46%; Fig. 5A). The final model comparison was performed to identify the most likely model with bilinear modulations. In accord with hypothesized feedback projections, the winning model within this family contained a bilinear modulation of the connectivity from TP to FG (Fig. 5B1). This indicates that top-down information of context increased effective connectivity from TP to FG, an antecedent region within the ventral visual processing stream. This model was 87.7% (EP) more likely than all other 14 models with different bilinear modulations (Fig. 5A), even though the architecture of other models was more complex.

BMS for the left hemisphere revealed similar results. On the first family level, the nonlinear families strikingly outperformed the bilinear family (family EP 99.55%; Fig. 5A). On the second family level, gating families including TP outperformed the gating families with a gating by aSTG only (EP 93.83%; Fig. 5A). On the third family level, BMS revealed the gating family of TP on FG to AMY as winning model family (EP 58.34%; Fig. 5A), followed by the second likely model family with a gating of TP and aSTG on fusiform to amygdalar coupling (EP 27.10%; Fig. 5A). The superiority of the gating family containing the TP as the only gating region was surprising given that in the previous study, the aSTG gating based on visual and auditory information alone reached similar EPs in left and right hemispheres (Pehrs et al. 2014). This might be due to the experimental design of the present study, in which perceptual information were combined with additional contextual information. Consistent with the right hemisphere

results, the final model comparison corroborated top-down modulations of TP to the ventral visual stream by showing that the model with bilinear modulation of TP to FG connectivity outperformed the other 14 models with an EP of 86.78% (Fig. 5B2). To note, the same model convincingly outperformed the other 14 models in the second winning model family containing a gating of both STG and TP (EP 98.84%).

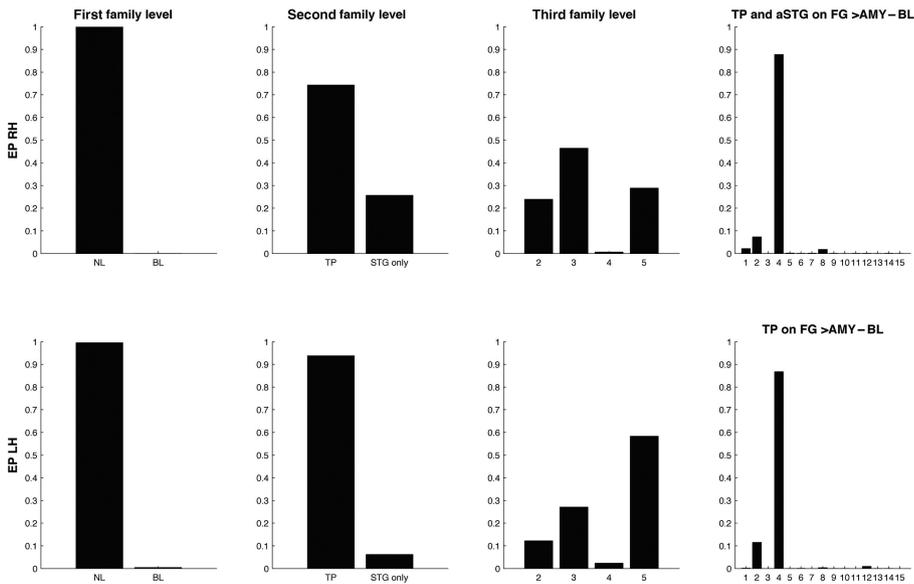
With regard to inference on model parameters, we focused on the experimental manipulations, namely the bilinear modulation of TP to FG connectivity, the nonlinear modulations, and the driving inputs of the winning model for both hemispheres (Tables 3 and 4). For model parameters of intrinsic connectivity, we focused on the 2 relevant connections from TP to FG and FG to AMY. For the right hemisphere, the nonlinear modulations of the aSTG and the TP were both suppressive on the connection from FG to AMY as indicated by negative signs of the parameters' group mean, which reached significance only for the aSTG [$t_{(25)} = 2.35$, $P = 0.027$; TP: $t_{(25)} = 0.90$, $P = 0.37$; Table 3 and Fig. 5C2]. Therefore, we were able to replicate the suppressive gating function of the aSTG on fusiform–amygdalar connectivity in the right hemisphere as reported in Pehrs et al. (2014) using audiovisual stimulus material of another emotion (sadness compared with happiness).

ANOVA for the bilinear modulations (sad context, film + music; neutral context, film + music; sad context, film + no music; neutral context, film + no) revealed a significant main effect of music ($F_{1,25} = 4.46$, $P < 0.05$), no significant main effect of context ($F_{1,25} = 0.001$, $P = 0.97$), and no music by context interaction ($F_{1,25} = 1.85$, $P = 0.18$; Table 3). This indicates that TP to FG connectivity is significantly enhanced by music. It is important to note that modulatory parameters reflect updates of endogenous connectivity based on experimental perturbations, here context and music. Consequently, the parameter estimates for intrinsic connection and modulatory effects are additive, resulting in 0.34 + 0.11 Hz (sad) and 0.34 + 0.22 Hz (neutral) or 0.34–0.08 Hz (sad) and 0.34–0.18 Hz (neutral) in RFX analysis. Therefore, the connectivity from TP to FG increased to 0.45 Hz (sad) and to 0.56 Hz (neutral) due to the auditory context conditions and decreased to 0.26 Hz (sad) and to 0.16 Hz (neutral) due to the mute context conditions. That all additive effects remain positive indicates that enhancing top-down modulations from TP to FG by context information are additionally increased by the music presented during the film clips. Inference on parameter estimates of driving inputs reached significant results for visual input to FG and auditory input to aSTG ($P < 0.001$). Beyond that, contextually framed film clips directly drive neurons in the TP (all $P < 0.05$, Table 3), supporting the strong sensitivity of TP for the processing of top-down semantic information (context).

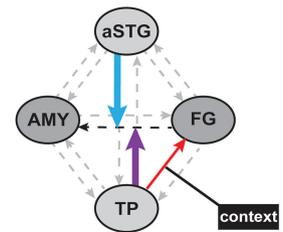
Consistent with the right hemisphere, the gating of TP on fusiform–amygdalar connectivity did not reach significance in the left hemisphere [$t_{(25)} = 0.34$, $P = 0.73$, Table 4 and Fig. 5C1], indicating that the direction and size of this gating does not have enough consistency across subjects and should be examined in future studies. Moreover, gating effects are not easy to interpret because they can occur directly on other connections or may be neurobiologically implemented by encompassing interneurons of neighboring neuronal populations. The suppressive nature of the TP and aSTG gating on fusiform–amygdalar effective connectivity may be due to several factors including the choice of the task design and the naturalistic stimulus material or reduced stimulus ambiguity during multisensory processing (cf. Pehrs et al. 2014).

ANOVA for the bilinear modulations (sad context, film + music; neutral context, film + music; sad context, film + no music;

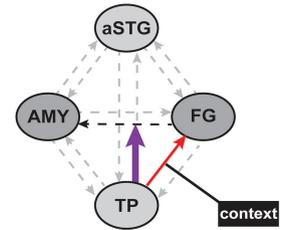
A BMS (RFX)



B1 RH optimal model (4)

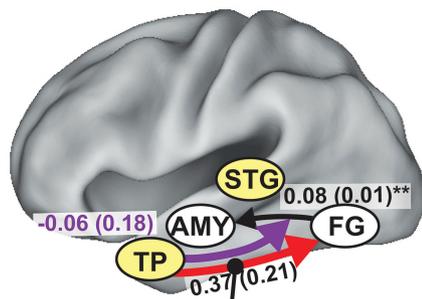


B2 LH optimal model (4)



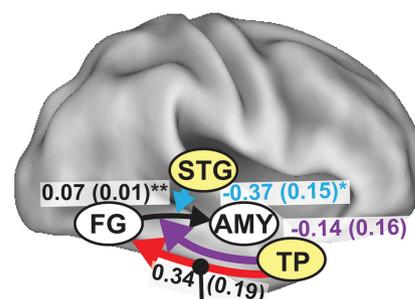
C BMS parameter analysis

C1 LH optimal model (4)



Modulatory input	
Sad context, film + music	0.12 (0.15)
Neutral context, film + music	0.003(0.15)
Sad context, film + no music	-0.31 (0.13)*
Neutral context, film + no music	-0.29 (0.14)

C2 RH optimal model (4)



Modulatory input	
Sad context, film + music	0.11 (0.12)
Neutral context, film + music	0.22 (0.10)*
Sad context, film + no music	-0.08 (0.17)
Neutral context, film + no music	-0.18 (0.10)

Figure 5. BMS, winning model and parameter analysis. (A) BMS using RFX inference on 4 family levels with EPs for the right hemisphere (RH) (upper panel) and left hemisphere (LH) (lower panel). From left to right: first family level: nonlinear (NL) versus bilinear families (BL); second family level: 3 families including TP gating versus family without TP gating (aSTG only); third family level: comparison of 4 remaining families with a gating of the aSTG on FG to AMY and TP on AMY to FG (Fig. 4C2), STG and TP on FG to AMY (Fig. 4C3), TP on AMY to FG (Fig. 4C4), and TP on FG to AMY (Fig. 4C5); fourth family level: comparison of bilinear models (BL) within the winning model family (gating of aSTG and TP on fusiform-amygdalar connectivity for the right hemisphere and gating of TP on fusiform-amygdalar connectivity for the left hemisphere) revealing model 4 as the optimal model. (B1) Winning model structure in the right hemisphere. (B2) Winning model structure in the left hemisphere. (C) Mean parameter estimates indicating connectivity strength across participants (SEM) in Hz. (C1) Left hemisphere with a gating of TP (purple arrow) and right hemisphere (C2) with a gating of TP and aSTG (purple arrow and light blue arrow) on FG to AMY connectivity. The parameter estimates of the endogenous connection from FG to AMY (black arrow) and TP to FG (red arrow) as well as the modulatory input of the 4 context conditions on TP to FG connectivity are listed (boxes below). Additive effects of modulatory parameters (boxes) on TP to FG intrinsic connectivity show that all context conditions increase effective connectivity from TP to FG in both hemispheres (for details refer to the Results section). ANOVA with parameter estimates of the modulatory input revealed a significant main effect of music (blue background), showing that top-down modulations of TP are increased by music. Regions converging multisensory stimuli are depicted in yellow. * $P < 0.05$, ** $P < 0.001$.

and neutral context, film + no) revealed a significant main effect of music ($F_{1,25} = 6.30$, $P < 0.05$), no significant main effect of context ($F_{1,25} = 0.18$, $P = 0.67$), and no music by context interaction

($F_{1,25} = 0.62$, $P = 0.43$; Table 4). In correspondence with the right hemisphere results, inference on parameter estimates shows that TP to FG connectivity is significantly enhanced by music.

Table 3 DCM parameters for the winning model in the right hemisphere

	Mean (SE)	One-sample t-test	
		T-value	P-value
Intrinsic connectivity			
TP ⇒ FG	0.34 (0.19)	1.76	0.088
FG ⇒ AMY	0.07 (0.01)	5.21	<0.001**
Driving input			
Visual on FG	0.25 (0.01)	12.97	<0.001**
Auditory on aSTG	0.67 (0.05)	12.37	<0.001**
Sad context, film + music on TP	0.03 (0.01)	3.93	0.001*
Neutral context, film + music on TP	0.02 (0.01)	3.91	0.001*
Sad context, film + no music on TP	0.02 (0.01)	3.71	0.001*
Neutral context, film + no music on TP	0.02 (0.01)	3.13	0.004*
Modulatory effects			
Sad context, film + music on TP ⇒ FG	0.11 (0.12)	0.90	0.37
Neutral context, film + music on TP ⇒ FG	0.22 (0.10)	2.18	0.03*
Sad context, film + no music on TP ⇒ FG	-0.08 (0.17)	0.48	0.63
Neutral context, film + no music on TP ⇒ FG	-0.18 (0.11)	1.71	0.09
Nonlinear modulations			
aSTG on FG ⇒ Amygdala	-0.37 (0.15)	2.35	0.02*
TP on FG ⇒ Amygdala	-0.14 (0.16)	0.90	0.37
ANOVA for modulatory effects			
Main effect plot		F-value	
Main effect music		0.001	0.97
Music by plot interaction		4.46	0.04*
		1.85	0.18

Note: Mean connectivity parameters with standard errors in brackets for model 4 (Fig. 4C3,B4) with modulatory effects of context on the connection from TP to FG and nonlinear effects by aSTG and TP on FG to AMY. t-tests and an ANOVA for modulatory effects were performed and T, F, and P-values (Bonferroni-corrected) are reported.

*P < 0.05.

**P < 0.001.

Table 4 DCM parameters for the winning model in the left hemisphere

	Mean (SE)	One-sample t-test	
		T-value	P-value
Intrinsic connectivity			
TP ⇒ FG	0.37 (0.21)	1.82	0.081
FG ⇒ AMY	0.08 (0.01)	9.47	<0.001**
Driving input			
Visual on FG	0.23 (0.01)	14.71	<0.001**
Auditory on aSTG	0.64 (0.05)	13.34	<0.001**
Sad context, film + music on TP	0.02 (0.01)	2.62	0.015*
Neutral context, film + music on TP	0.02 (0.01)	2.96	0.007*
Sad context, film + no music on TP	0.01 (0.01)	2.27	0.032*
Neutral context, film + no music on TP	0.01 (0.01)	1.18	0.248
Modulatory effects			
Sad context, film + music on TP ⇒ FG	0.12 (0.15)	0.79	0.43
Neutral context, film + music on TP ⇒ FG	0.003 (0.15)	0.02	0.98
Sad context, film + no music on TP ⇒ FG	-0.31 (0.13)	2.26	0.03*
Neutral context, film + no music on TP ⇒ FG	-0.29 (0.14)	2.02	0.05
Nonlinear modulations			
TP on FG ⇒ Amygdala	-0.06 (0.18)	0.34	0.73
ANOVA for modulatory effects			
Main effect plot		F-value	
Main effect music		0.18	0.67
Music by plot interaction		6.30	0.01*
		0.62	0.43

Note: Mean connectivity parameters with standard errors in brackets for model 4 (Fig. 4C5,B4) with modulatory effects of context on the connection from TP to FG and the nonlinear effect by TP on FG to AMY. t-tests and an ANOVA for modulatory effects were performed and T, F, and P-values (Bonferroni-corrected) are reported.

*P < 0.05.

**P < 0.001.

The connectivity from TP to FG increased to 0.49 Hz (sad) and to 0.37 Hz (neutral) due to the auditory context conditions and decreases to 0.06 Hz (sad) and to 0.08 Hz (neutral) due to the mute

context conditions. Notably, also in the left hemisphere, all additive effects remain positive, indicating that all context conditions increase TP to FG effective connectivity within this experiment.

Inference on parameter estimates of driving inputs reached highly significant results for visual input to FG and auditory input to aSTG ($P < 0.001$). Contextually framed film clips directly drive neurons in the TP, except for the condition, neutral context, and film + no music (all $P < 0.05$, Table 4).

Discussion

The aim of this study was to examine the role of TP in social cognition. Building on the semantic hub theory and Damasio's theory of heteromodal convergence regions, we asked whether TP integrates visual, auditory, and contextual information and how this interacts with emotion processing, specifically with empathic concern for movie characters. By demonstrating an emotion effect of context, the present study places TP at the core of social cognition with a critical role in modulating the activation of other brain areas. In line with our hypothesis, BMS revealed that both TP and aSTG gate the connectivity from FG to AMY in a suppressive manner. Moreover, contextual information had a strengthening effect on the effective connectivity from TP to FG during movie perception, demonstrating top-down modulations within the ventral visual stream. Crucially, inspection on parameter estimates revealed that these backward projections were enhanced by a simultaneous presentation of emotion-congruent music. Consistent with studies on object recognition (Tyler et al. 2004; Chan et al. 2011), we demonstrate increased top-down modulations from TP within the ventral stream as a function of complexity and increasing multisensory integration demands. To the best of our knowledge, this is the first study to show top-down modulations of TP in the context of social cognition. We discuss these modulations with top-down facilitation models (Bar 2003; Bar et al. 2006) and infer general implications for social cognition.

Emotion Effects of Sad Context in the TP

Using highly emotional audiovisual material (sad film clips combined with sad music) and variations of sad and neutral contextual framings, the sensitivity of TP for socioemotional processing was investigated (Zahn et al. 2007; Wagner et al. 2011; Olson et al. 2013; Hassabis et al. 2014). Enhanced TP activation was found when participants were previously informed about the tragic circumstances leading to the protagonists' situation. The same anatomical location has been reported in previous studies showing emotion effects of sad and fearful contexts for the processing of neutral social scenes and faces (Mobbs et al. 2006; Willems et al. 2011).

The parametric analysis revealed that activations in the right TP and STG linearly increased with the participants' empathic concern for the movie characters, consistent with positive correlations between empathy scores and volume of the right TP (Ran-kin et al. 2006), and compatible with the present emotion effect of sad context.

Both analyses, the comparison of sad against neutral context conditions and the parametric analysis, revealed TP activations with regional overlap. However, the cluster responsive to empathic concern was located more medially and spread more posteriorly, whereas the cluster showing an emotion effect of sad context was located more anteriorly and laterally (Fig. 3 and Supplementary Fig. 2). In most studies, the TP is treated as a homogenous region (BA38) with a unified function. Contradicting this view, a recent study has shown a connectivity-based parcellation of the TP in 3 distinct areas, which appear to be key for different processes (Fan et al. 2014). In line with this segregation, the emotion effect of sad context corresponds to the medial TP

subregion, which is characterized by connections with emotion-related areas (Fan et al. 2014). Empathic concern, however, rather involves the lateral TP subregion, which shows strong positive correlations with regions of the default mode network (Fan et al. 2014), generally discussed to be involved in mentalizing and empathy (Schilbach et al. 2008; Mars et al. 2012). Therefore, the present anatomical dissociation of TP activation might indicate functionally specialized subregions as described by Fan et al. (2014).

Activation in the right TP has repeatedly been associated with better performance in theory of mind and empathy tasks (Völlm et al. 2006; Mier et al. 2010). Correspondingly, TP is a well-established region within the mentalizing network and considered to retrieve socioemotional knowledge from episodic memory (Frith and Frith 2003). Of note is the particular involvement of TP in processing of negative emotions, predominantly sadness (Blair et al. 1999; Levesque et al. 2003; Omar et al. 2011; Downey et al. 2013), and adverse aspects of empathy (Jimura et al. 2010; Lenzi et al. 2013; Seehausen et al. 2014). Therefore, we join Dolan et al. (2000) and Moriguchi et al. (2006) in asserting that empathic concern might be mediated by recall of prior negative experiences [see also Perry et al. (2011)]. Future studies are needed to examine how activity in TP interacts with autobiographic episodic memory and how this affects the modulatory impact of TP on other brain regions.

Top-Down Facilitation in the Ventral Visual Pathway

Our DCM results revealed robust evidence for the neuromodulatory role of TP in social cognition. In addition to a suppressing gating effect of TP on fusiform-amygdalar connectivity, context-dependent modulations of TP-FG coupling were shown, such that TP top-down modulates antecedent regions of the ventral visual stream. The ventral visual stream is implicated in mapping perceptual information onto meaning (Ungerleider and Haxby 1994). In principle, visual stimuli evoke neural activity propagating through the ventral stream from lower-level visual areas (V1, V2, and V4) over the FG (Gross 2008) peaking in the TP, thereby continuously increasing levels of processing complexity (Ungerleider and Mishkin 1982; Felleman and Van Essen 1991). An intriguing study recently supported this functional-anatomical dissociation. Using multivoxel pattern analysis, Peelen and Caramazza (2012) revealed posterior areas to process perceptual properties of objects, whereas conceptual properties (e.g., how an object is used) were processed more anteriorly in the ventral temporal lobe. However, the question of how and under what circumstances neural interactions between different levels of processing complexity occur still remains unresolved.

Top-down facilitation models were developed in the context of object recognition and account for feedback connections in the ventral visual pathway (Ullman 1995; Bar 2003; Simmons and Barsalou 2003; Bar et al. 2006). When, for example, visual information is coarse and ambiguous, activity in higher-level regions, such as prefrontal cortex or aTL, generate a set of possible higher-order representations (i.e., "initial guesses," Bar 2003) that modulate processing in "lower-level" structures (Miller and Cohen 2001). Their integration with the bottom-up process via recurrent connections facilitates object recognition by top-down modulations of lower-level analysis.

Facilitation models are neuroanatomically supported by demonstrations of recurrent pathways in macaque monkeys (Gilbert and Li 2013; Kravitz et al. 2013) and functionally supported by studies showing increased feedback projections from

prefrontal and anterior temporal regions to posterior temporal areas with increasing semantic integration demands during object recognition (Tyler et al. 2004; Chan et al. 2011; Clarke et al. 2011; Yvert et al. 2012). In these studies, integration demands refer to rising complexity like word recognition compared with phoneme detection and object recognition on different levels of specificity (basic vs. domain). A lesion study by Campo et al. (2013) corroborates the importance of top-down modulations from TP by showing that backward projections during naming of pictures are reduced in patients with atrophy in aTL.

Equivalent to rising top-down modulations with increasing integration demands, backward projections from TP to FG were enhanced here by co-presentation of emotion-congruent music. Besides written context information preceding the film clips, the music additionally provides auditory context information during the presentation of the film clips by forwarding musical meaning emerging, for example, from emotional expression. Music has a powerful capacity to express and elicit emotions (Blood and Zatorre 2001; Koelsch 2013), to modulate emotion processing of dynamic visual stimuli (Eldar et al. 2007; Pehrs et al. 2014), and to trigger mentalizing processes (Steinbeis and Koelsch 2009; Parsons et al. 2014). Accordingly, patients with semantic dementia show deficits in recognizing musical emotions and are impaired in attributing affective mental associations with musical stimuli to the degree to which they show atrophy in the aTL (Omar et al. 2011; Hsieh et al. 2012; Downey et al. 2013). Therefore, music increases task demands and top-down modulations by providing (1) perceptual information from another modality and (2) musical context information during the presentation of the film clips that need to be integrated with the visual information and the written context information preceding the film clips. Increased backward projections by music also reinforce the role of TP to act as a semantic hub integrating multimodal information (Patterson et al. 2007), a capacity that according to Miller and Cohen (2001) is necessary for a region to perform top-down modulations via feedback signals.

Our results synthesize Damasio's concept of convergence regions with top-down facilitation models. According to Damasio (1989), binding of unisensory information is performed in convergence regions, which operate at different levels of complexity. While the aSTG binds features on a perceptual level, the TP is involved in higher-order conceptual processing (Damasio 1989). TP receives information from modality-specific cortices and assigns semantic knowledge across modalities (Patterson et al. 2007; Visser et al. 2010), whereas the STG performs early audiovisual integration within modality-specific cortices (Kreifelts et al. 2007). Both Damasio's theory of retroactivation (1989) and the semantic hub accounts assume that semantic knowledge is processed in the aTL (McClelland and Rogers 2003; Patterson et al. 2007).

Here, we use DCM to connect this idea with top-down facilitation models (Bar 2003) by testing top-down modulations of TP to FG as a function of increasing task demands by context information and multisensory (audiovisual) versus unisensory (visual) stimulation. We show that convergence regions (aSTG and TP) modulate fusiform-amygdalar connectivity. We also show that top-down modulations from TP to lower-level perceptual areas (i.e., FG) increase with context information and with multisensory integration demands (Fig. 5C). We thus provide evidence for the TP to act as a convergence region and a top-down node at the same time. The present study supports TP's role in integrating multisensory information into a meaningful representation. This representation is, in turn, provided to other processing streams as indicated by TP's neuromodulatory impact. Our results

broaden the concept of convergence regions and semantic hub accounts with top-down facilitation models that were transferred here from object recognition to higher-order social cognition.

A similar approach is provided by current models of cue integration (Zaki 2013) that conceive social cognition as an integrative process over multimodal cues. In line with Bayesian approaches to physical perception (Ernst and Bulthoff 2004; Friston and Stephan 2007), we found that the more social cues converge over modalities, that is, the stronger conditional probabilities of the character being in a miserable state (i.e., the more free energy is minimized, Friston 2010), the stronger the empathic concern of the participants, and top-down modulations of TP. We thus translated a basic principle of physical perception (object recognition) to complex social cognition and account for the plausibility of cue integration as a tool to develop Bayesian models of social cognition (Zaki 2013).

According to Bar (2003), the extent and dynamics of top-down modulations can be modulated by task demands and context information (e.g., ambiguous input, priming, and surrounding objects). This may be specifically important for complex real-life social scenarios, characterized by constantly changing perceptual input with rapid changes of integration demands. Whether top-down modulations are suppressing or facilitating might fluctuate as rapid as the perceptual input. Multimodal imaging using methods with a higher temporal resolution, for example magnetoencephalography (MEG), could elucidate possible bi-directional bottom-up and top-down interactions in future research.

Limitations

Although our study provides a neurobiological mechanism for the role of TP in social cognition, several limitations should be noted. First, results from univariate analyses revealed a bilateral cluster in the TP when protagonists in empathic film clips were embedded in a sad compared with a neutral context. While this result was expected from previous studies, the absence of significant clusters beyond TP should be interpreted with caution. As indicated by prior behavioral rating procedures, all characters were perceived as clearly sad, which could have automatically induced empathic concern. To control for this effect, neutral visual stimuli would have been advantageous, but were not used in the present study to avoid possible incongruity effects (sad context and neutral face).

Second, the presentation of context before the empathic film clips revealed BOLD activity beyond TP, for example, in the TPJ, which is also part of the human mentalizing network (Van Overwalle 2009). Questions addressing interactions within the mentalizing network including different nodes in the network for DCM analysis were thus feasible. However, the present network structure was established previously in studies on multisensory integration using similar naturalistic stimuli from feature films (Pehrs et al. 2014), while it is also determined by the aim of the present study to examine interactions of multisensory integration and emotion processing focusing on the modulatory impact of TP.

Finally, our results are based on empathic film clips of sadness and would need to be replicated for generalizability to other emotions.

Conclusion and Outlook

The present study revealed the TP as a key structure for empathic concern and supports its role in socioemotional cognition (Olson et al. 2007, 2013). More importantly, the findings offer a new conception of the TP by pointing to its neuromodulatory role.

We show top-down modulations from complex processing in TP to lower-level perceptual processing in the FG when integration involves narrative context. These results challenge the traditional view of a strictly feed-forward processing along the ventral visual stream. Instead, they indicate interactive neural propagation within the stream consistent with findings from object recognition. In line with the amodal semantic hub theory, we argue that during social cognition, TP integrates visual, auditory, and semantic emotional information and that top-down facilitation is increased as a function of integration demands by prior expectations based on contextual knowledge and information from different modalities (here music). By using systematic changes of naturalistic stimuli together with DCM, this study provides a novel approach to social cognition research in balancing experimental control and naturalistic perception. Prospective studies should investigate whether TP not only exerts top-down modulations but also provides bottom-up perceptual information to higher-order cognitive areas, as indicated, for example, by functional connectivity to the orbital prefrontal cortex (Kahnt et al. 2012).

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Funding

This work was supported by Excellence Initiative of the German Federal Ministry of Education and Research Deutsche Forschungsgemeinschaft Grant EXC 302 and from a project of the German Federal Ministry of Education and Research (JA 823/4-2).

Notes

The authors thank Mareike Voget for help with stimulus material and data acquisition and Sabine Aust for valuable comments on a previous version of the manuscript, as well as Felix Blankenburg and Ryszard Auksztulewicz for fruitful discussions. *Conflict of Interest*: None declared.

References

- Altmann U, Bohrn IC, Lubrich O, Menninghaus W, Jacobs AM. 2012. The power of emotional valence—from cognitive to affective processes in reading. *Front Hum Neurosci.* 6:192.
- Aust S, Alkan Härtwig E, Koelsch S, Heekeren HR, Heuser I, Bajbouj M. 2013. How emotional abilities modulate the influence of early life stress on hippocampal functioning. *Soc Cogn Affect Neurosci.* 9:1038–1045.
- Bar M. 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. *J Cogn Neurosci.* 15:600–609.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, et al. 2006. Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA.* 103:449–454.
- Baumgartner T, Lutz K, Schmidt CF, Jäncke L. 2006. The emotional power of music: how music enhances the feeling of affective pictures. *Brain Res.* 1075:151–164.
- Binney RJ, Embleton KV, Jefferies E, Parker GJM, Lambon Ralph MA. 2010. The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cereb Cortex.* 20:2728–2738.
- Blair RJ, Morris JS, Frith CD, Perret DI, Dolan RJ. 1999. Dissociable neural responses to facial expressions of sadness and anger. *Brain.* 122:883–893.
- Blood AJ, Zatorre RJ. 2001. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci USA.* 98:11818–11823.
- Campo P, Poch C, Toledano R, Igoa JM, Belincho M, Garcia-Morales I, Gil-Nagel A. 2013. Anterobasal temporal lobe lesions alter recurrent functional connectivity within the ventral pathway during naming. *J Neurosci.* 33:12679–12688.
- Carlson TA, Simmons RA, Kriegeskorte N, Slevc LR. 2014. The emergence of semantic meaning in the ventral temporal pathway. *J Cogn Neurosci.* 26:120–131.
- Catani M, Jones DK, Donato R, Ffytche DH. 2003. Occipito-temporal connections in the human brain. *Brain.* 126:2093–2107.
- Chan AM, Baker JM, Eskandar E, Schomer D, Ulbert I, Marinkovic K, Cash SS, Halgren E. 2011. First-pass selectivity for semantic categories in human anteroventral temporal lobe. *J Neurosci.* 31:18119–18129.
- Clarke A, Taylor KI, Tyler LK. 2011. The evolution of meaning: spatio-temporal dynamics of visual object recognition. *J Cogn Neurosci.* 23:1887–1899.
- Crosby EC. 1963. Correlative anatomy of the nervous system. *Acad Med.* 38:526.
- Damasio AR. 1989. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition.* 33:25–62.
- Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Moss HE, Matthews PM, Tyler LK. 2000. Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. *Neuroimage.* 11:589–600.
- Dolan RJ, Lane R, Chua P, Fletcher P. 2000. Dissociable temporal lobe activations during emotional episodic memory retrieval. *Neuroimage.* 11:203–209.
- Dolan RJ, Morris JS, de Gelder G. 2001. Crossmodal binding of fear in voice and face. *Proc Natl Acad Sci USA.* 98:10006–10010.
- Downey LE, Blezat A, Nicholas J, Omar R, Golden HL, Mahoney CJ, Crutch SJ, Warren JD. 2013. Mentalising music in frontotemporal dementia. *Cortex.* 49:1844–1855.
- Eldar E, Ganor O, Admon R, Bleich A, Hendler T. 2007. Feeling the real world: limbic response to music depends on related content. *Cereb Cortex.* 17:2828–2840.
- Ernst MO, Bulthoff HH. 2004. Merging the senses into a robust percept. *Trends Cogn Sci.* 8:162–169.
- Fan L, Wang J, Zhang Y, Han W, Yu C, Jiang T. 2014. Connectivity-based parcellation of the human temporal pole using diffusion tensor imaging. *Cereb Cortex.* 24:3365–3378.
- Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex.* 1:1–47.
- Friston KJ. 2010. The free-energy principle: a unified brain theory? *Nat Rev Neurosci.* 11:127–138.
- Friston KJ, Harrison L, Penny W. 2003. Dynamic causal modelling. *Neuroimage.* 19:1273–1302.
- Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. 1995. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp.* 2:189–210.
- Friston KJ, Mechelli A, Turner R, Price CJ. 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage.* 12:466–477.
- Friston KJ, Stephan KE. 2007. Free-energy and the brain. *Synthese.* 159:417–458.

- Frith U, Frith CD. 2003. Development and neurophysiology of mentalizing. *Philos Trans R Soc Lond B Biol Sci.* 358:459–473.
- Gilbert CD, Li W. 2013. Top-down influences on visual processing. *Nat Rev Neurosci.* 14:350–363.
- Goetz JL, Keltner D, Simon-Thomas E. 2010. Compassion: an evolutionary analysis and empirical review. *Psychol Bull.* 136:351–374.
- Gross CG. 2008. Single neuron studies of inferior temporal cortex. *Neuropsychologia.* 46:841–852.
- Hassabis D, Spreng RN, Rusu AA, Robbins CA, Mar RA, Schacter DL. 2014. Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cereb Cortex.* 24:1979–1987.
- Hickok G, Poeppel D. 2007. The cortical organization of speech processing. *Nat Rev Neurosci.* 8:393–402.
- Hickok G, Poeppel D. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition.* 92:67–99.
- Hsieh S, Hornberger M, Piguet O, Hodges JR. 2012. Brain correlates of musical and facial emotion recognition: evidence from the dementias. *Neuropsychologia.* 50:1814–1822.
- Jeannerod M, Jacob P. 2005. Visual cognition: a new look at the two visual systems model. *Neuropsychologia.* 43:301–312.
- Jeong JW, Diwadkar VA, Chugani CD, Sinsonngsud P, Muszik O, Behen ME, Chugani HT. 2011. Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation. *Neuroimage.* 54:2973–2982.
- Jimura K, Konishi S, Miyashita Y. 2009. Temporal pole activity during perception of sad faces, but not happy faces, correlates with neuroticism trait. *Neurosci Lett.* 453:45–48.
- Jimura K, Konishi S, Asari T, Miyashita Y. 2010. Temporal pole activity during understanding other persons' mental states correlates with neuroticism trait. *Brain Res.* 1328:104–112.
- Kahnt T, Chang LJ, Park SQ, Heinze J, Haynes JD. 2012. Connectivity-based parcellation of the human orbitofrontal cortex. *J Neurosci.* 32:6240–6250.
- Koelsch S. 2013. Brain-correlates of music-evoked emotions. *Nat Rev Neurosci.* 15:170–180.
- Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci.* 17:26–49.
- Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D. 2007. Audio-visual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage.* 37:1445–1456.
- Lambon Ralph MA, Ehsan S, Baker GA, Rogers TT. 2012. Semantic memory is impaired in patients with unilateral anterior temporal lobe resection for temporal lobe epilepsy. *Brain.* 135:242–258.
- Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT. 2000. Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp.* 10:120–131.
- Lenzi D, Trentini C, Pantano P, Macaluso E, Lenzi GL, Ammaniti M. 2013. Attachment models affect brain responses in areas related to emotions and empathy in nulliparous women. *Hum Brain Mapp.* 34:1399–1414.
- Levesque J, Eugene F, Joanette Y, Paquette V, Mensour B, Beaudoin G, Leroux JM, Bourgouin P, Beauregard M. 2003. Neural circuitry underlying voluntary suppression of sadness. *Biol Psychiatry.* 53:502–510.
- Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage.* 19:1233–1239.
- Manera V, Samson AC, Pehrs C, Lee IA, Gross JJ. 2014. The eyes have it: the role of attention in cognitive reappraisal of social stimuli. *Emotion.* 14:833–839.
- Mars RB, Neubert FX, Noonan MP, Sallet J, Toni I, Rushworth MFS. 2012. On the relationship between the “default mode network” and the “social brain”. *Front Hum Neurosci.* 6:189.
- McClelland JL, Rogers TT. 2003. The parallel distributed processing approach to semantic cognition. *Nat Rev Neurosci.* 4:310–322.
- Mier D, Lis S, Neuthe K, Sauer C, Esslinger C, Gallhofer B, Kirsch P. 2010. The involvement of emotion recognition in affective theory of mind. *Psychophysiology.* 47:1028–1039.
- Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. *Ann Rev Neurosci.* 24:167–202.
- Mishkin M, Ungerleider LG, Macko KA. 1983. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6:414–417.
- Mobbs D, Weiskopf N, Lau HC, Featherstone E, Dolan RJ, Frith CD. 2006. The Kuleshov effect: the influence of contextual framing on emotional attributions. *Soc Cogn Affect Neurosci.* 1:95–106.
- Moriguchi Y, Ohnishi T, Lane RD, Maeda M, Mori T, Nemoto K, Matsuda H, Komaki G. 2006. Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. *Neuroimage.* 32:1472–1482.
- Müller VI, Cieslik EC, Turetsky BI, Eickhoff SB. 2012. Crossmodal interactions in audiovisual emotion processing. *Neuroimage.* 60:553–561.
- Nakamura K, Kubota K. 1996. The primate temporal pole: its putative role in object recognition and memory. *Behav Brain Res.* 77:53–77.
- Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia.* 9:97–113.
- Olson IR, Plotzker A, Ezzyat Y. 2007. The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain.* 130:1718–1731.
- Olson IR, McCoy D, Klobusicky E, Ross LA. 2013. Social cognition and the anterior temporal lobes: a review and theoretical framework. *Soc Cogn Affect Neurosci.* 8:123–133.
- Omar R, Henley SM, Bartlett JW, Hailstone JC, Gordon E, Sauter DA, Frost C, Scott SK, Warren JD. 2011. The structural neuroanatomy of music emotion recognition: evidence from frontotemporal lobar degeneration. *Neuroimage.* 56:1814–1821.
- Parkinson C, Wheatley T. 2014. Relating anatomical and social connectivity: white matter microstructure predicts emotional empathy. *Cereb Cortex.* 24:614–625.
- Parsons CE, Young KS, Jegindø EM, Vuust P, Stein A, Kringelbach ML. 2014. Music training and empathy positively impacts adults' sensitivity to infant distress. *Front Psychol.* 5:1440.
- Patterson K, Nester PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci.* 8:976–987.
- Peelen MV, Caramazza A. 2012. Conceptual object representations in human anterior temporal cortex. *J Neurosci.* 32:15728–15736.
- Pehrs C, Deserno L, Bakels JH, Schlochtermeyer LH, Kappelhoff H, Jacobs AM, Fritz T, Koelsch S, Kuchinke L. 2014. How music alters a kiss: superior temporal gyrus controls fusiform-amygdala effective connectivity. *Soc Cogn Affect Neurosci.* 9:1770–1778.
- Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP. 2010. Comparing families of dynamic causal models. *PLoS Comput Biol.* 6:e1000709.

- Perry D, Hendler T, Shamay-Tsoory SG. 2011. Projecting memories: the role of the hippocampus in emotional mentalizing. *Neuroimage*. 54:1669–1676.
- Plantinga C. 1999. The scene of empathy and the human face on film. In: Plantinga C, Smith GM, editors. *Passionate views. Film, cognition and emotion*. Baltimore: London. p. 239–255.
- Rankin KP, Gorno-Tempini ML, Allison SC, Stanley CM, Glenn S, Weiner MW, Miller BL. 2006. Structural anatomy of empathy in neurodegenerative disease. *Brain*. 129:2945–2956.
- Raz G, Jacob Y, Gonen T, Winetraub Y, Flash T, Soreq E, Hendler T. 2014. Cry for her or cry with her: context-dependent dissociation of two modes of cinematic empathy in network cohesion dynamics. *Soc Cogn Affect Neurosci*. 9:30–38.
- Royet JP, Zald D, Versace R, Costes N, Lavenne F, Koenig O, Gervais R. 2000. Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: a positron emission tomography study. *J Neurosci*. 20:7752–7759.
- Schilbach L, Eickhoff SB, Rotarska-Jagiela A, Fink GR, Vogeley K. 2008. Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the “default system” of the brain. *Conscious Cogn*. 17:457–467.
- Seehausen M, Kazzner P, Bajbouj M, Heekeren HR, Jacobs AM, Klann-Delius G, Menninghaus W, Prehn K. 2014. Talking about social conflict in the MRI scanner: neural correlates of being empathized with. *Neuroimage*. 84:951–961.
- Simmons WK, Barsalou LW. 2003. The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn Neuropsychol*. 20:451–486.
- Simmons WK, Reddish M, Bellgowan PSF, Martin A. 2010. The selectivity and functional connectivity of the anterior temporal lobes. *Cereb Cortex*. 20:813–825.
- Skipper IM, Ross LA, Olson IR. 2011. Sensory and semantic category subdivisions within the anterior temporal lobes. *Neuropsychologia*. 49:3419–3429.
- Stefanacci L, Amaral DG. 2002. Some observations on cortical inputs to the macaque monkey amygdala: an anterograde tracing study. *J Comp Neurol*. 451:301–323.
- Steinbeis N, Koelsch S. 2009. Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. *Cereb Cortex*. 19:619–623.
- Stephan KE, Kasper L, Harrison LM, Daunizeau J, den Ouden HE, Breakspear M, Friston KJ. 2008. Nonlinear dynamic causal models for fMRI. *Neuroimage*. 42:649–662.
- Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. 2009. Bayesian model selection for group studies. *Neuroimage*. 46:1004–1017.
- Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ. 2007. Comparing hemodynamic models with DCM. *Neuroimage*. 38:387–401.
- Tyler LK, Stamatakis EA, Bright P, Acres K, Abdallah S, Rodd JM, Moss HE. 2004. Processing objects at different levels of specificity. *J Cogn Neurosci*. 16:351–362.
- Ullman S. 1995. Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb Cortex*. 1:1–11.
- Ungerleider LG, Haxby JV. 1994. “What” and “where” in the human brain. *Curr Opin Neurobiol*. 4:157–165.
- Ungerleider LG, Mishkin M. 1982. Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield RJW, editors. *Analysis of visual behavior*. Cambridge, MA: MIT Press. p. 549–586.
- Van Overwalle F. 2009. Social cognition and the brain: a meta-analysis. *Hum Brain Mapp*. 30:829–858.
- Visser M, Jefferies E, Lambon Ralph MA. 2010. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J Cogn Neurosci*. 22:1083–1094.
- Völlm BA, Taylor AN, Richardson P, Corcoran R, Stirling J, McKie S, Deakin JF, Elliot R. 2006. Neurofunctional correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage*. 29:90–98.
- Wagner DD, Kelley WM, Heatherton TF. 2011. Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cereb Cortex*. 21:2788–2796.
- Willems RM, Clevis K, Hagoort P. 2011. Add a picture for suspense: neural correlates of the interaction between languages and visual information in the perception of fear. *Soc Cogn Affect Neurosci*. 6:404–416.
- Yvert G, Perrone-Bertolotti M, Baciú M, David O. 2012. Dynamic causal modeling of spatiotemporal integration of phonological and semantic processes: an electroencephalographic study. *J Neurosci*. 32:4297–4306.
- Zahn R, Moll J, Iyengar V, Huey ED, Tierney M, Krueger F, Grafman J. 2009. Social conceptual impairments in frontotemporal lobar degeneration with right anterior temporal hypometabolism. *Brain*. 132:604–616.
- Zahn R, Moll J, Krueger F, Huey ED, Garrido G, Grafman J. 2007. Social concepts are represented in the superior anterior temporal cortex. *Proc Natl Acad Sci USA*. 104:6430–6435.
- Zaki J. 2013. Cue integration: a common framework for social cognition and physical perception. *Perspect Psychol Sci*. 8:296–312.
- Zaki J, Ochsner K. 2012. The neuroscience of empathy: progress, pitfalls and promise. *Nat Neurosci*. 15:675–680.