

RESEARCH ARTICLE

Supramodal neural networks support top-down processing of social signals

Melina Sonderfeld^{1,2} | Klaus Mathiak^{1,2} | Gianna S. Häring^{1,2} | Sarah Schmidt³ | Ute Habel^{1,2} | Raquel Gur⁴ | Martin Klasen^{1,2,5} 

¹Department of Psychiatry, Psychotherapy, and Psychosomatics, Medical School, RWTH Aachen, Aachen, Germany

²JARA-Translational Brain Medicine, RWTH Aachen University, Aachen, Germany

³Life & Brain - Institute for Experimental Epileptology and Cognition Research, Bonn, Germany

⁴Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

⁵Interdisciplinary Training Centre for Medical Education and Patient Safety - AIXTRA, Medical Faculty, RWTH Aachen University, Aachen, Germany

Correspondence

Martin Klasen, Interdisciplinary Training Centre for Medical Education and Patient Safety - AIXTRA, Medical Faculty, RWTH Aachen University, Forckenbeckstr. 71, 52074 Aachen, Germany.
Email: mklasen@ukaachen.de

Funding information

Bundesministerium für Bildung und Forschung, Grant/Award Number: APIC: 01EE1405B; Deutsche Forschungsgemeinschaft, Grant/Award Numbers: IRTG 1328, IRTG 2150; ICCR Aachen; Brain Imaging Facility of the Interdisciplinary Center for Clinical Research (ICCR)

Abstract

The perception of facial and vocal stimuli is driven by sensory input and cognitive top-down influences. Important top-down influences are attentional focus and supramodal social memory representations. The present study investigated the neural networks underlying these top-down processes and their role in social stimulus classification. In a neuroimaging study with 45 healthy participants, we employed a social adaptation of the Implicit Association Test. Attentional focus was modified via the classification task, which compared two domains of social perception (emotion and gender), using the exactly same stimulus set. Supramodal memory representations were addressed via congruency of the target categories for the classification of auditory and visual social stimuli (voices and faces). Functional magnetic resonance imaging identified attention-specific and supramodal networks. Emotion classification networks included bilateral anterior insula, pre-supplementary motor area, and right inferior frontal gyrus. They were pure attention-driven and independent from stimulus modality or congruency of the target concepts. No neural contribution of supramodal memory representations could be revealed for emotion classification. In contrast, gender classification relied on supramodal memory representations in rostral anterior cingulate and ventromedial prefrontal cortices. In summary, different domains of social perception involve different top-down processes which take place in clearly distinguishable neural networks.

KEYWORDS

emotion, gender, salience network, social perception, supramodal, top-down

1 | INTRODUCTION

The social evaluation of another person is based on visual and auditory signals, such as facial and vocal cues (Hensel, Bzdok, Müller, Zilles, & Eickhoff, 2015; Klasen, Chen, & Mathiak, 2012; Klasen, Kenworthy,

Mathiak, Kircher, & Mathiak, 2011; Massaro & Egan, 1996). Social perception encompasses different kinds of information about our counterpart, including variable conditions such as the current affective state, but also fixed characteristics such as gender (Joassin, Maurage, & Campanella, 2011), up to complex social judgments about traits (Adolphs, 2003). Thus, social perception has great influence on our behavior toward others (Alcalá-López et al., 2018; Bzdok et al., 2012;

Melina Sonderfeld and Klaus Mathiak authors contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

Hughes, Dispenza, & Gallup, 2004; Todorov, 2008). It is well established that the perception of auditory and visual social stimuli is driven not only by physical stimulus properties, but also by top-down processes (Gilbert & Li, 2013; Latinus, VanRullen, & Taylor, 2010). "Top-down processes" is a collective term for various types of cognitive influences driving perception. Important top-down influences on perception are attentional focus, that is, the aspect of a stimulus that a person is attending to (Corbetta & Shulman, 2002; Hopfinger, Buonocore, & Mangun, 2000; van Atteveldt, Formisano, Goebel, & Blomert, 2007) or supramodal representations in long-term memory (Choi, Lee, & Lee, 2018; Ramsey, Cross, & Hamilton, 2013). Some previous studies have addressed the role of specific top-down contributions in social perception. Bzdok et al. (2012) separated neural networks underlying social, face-specific, emotional and cognitive stimulus processing aspects. These findings suggest that there are neural networks that are driven by top-down influences such as the task, but not by the stimulus material itself. Further evidence for this notion comes from a study by Hensel et al. (2015), who identified an involvement of dorsomedial prefrontal cortex (DLPFC) specifically during social trait judgments irrespective from stimulus modality.

Following the line of these studies, the present study investigated the neural networks underlying two types of top-down influences on the perception of voices and faces: attentional focus (i.e., the attended aspect of the stimulus material) and memory representations. Attentional focus was varied via the task, that is, attending to either emotion or gender of the faces and voices. Moreover, social evaluation requires a comparison with a representation in the individual's long-term (or "reference") memory (Roitblat, 1987), which has been formed via previous experience (Mazur, 2017). For social evaluation, we were interested if the respective networks were supramodal, that is, independent from stimulus modality. To identify such supramodal memory representations, we developed the Social Implicit Association Test (SIAT), a social variant of the well-established Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The SIAT is described in detail in the Section 2; in short, it investigates associations between memory representations via reaction times to the respective stimuli. In the original IAT, associated stimuli (such as the words "doctor" and "hospital") lead to faster responses than non-associated stimuli (such as "bird" and "cigarette"; cf. Collins & Loftus, 1975). In the SIAT, we similarly assumed supramodal associations between the same social categories in voice and face, for example, between a happy face and a happy voice. From a neurobiological perspective, we assumed that such an association may be reflected by a shared brain region. In other words, we assumed that associated representations are in fact two aspects of one and the same concept and represented in the same brain region. With respect to faces and voices, this would correspond to a supramodal memory representation. There are previous functional magnetic resonance imaging (fMRI) studies on the IAT following the same logic. As an example, Knutson, Mah, Manly, and Grafman (2007) used the IAT during fMRI to investigate the neural substrates of gender and racial bias, identifying ventromedial prefrontal cortex (VMPFC) and ventral anterior cingulate cortex (vACC) as putative regions. These findings

are well in line with Milne and Grafman (2001), who found a reduced IAT effect in patients with VMPFC lesions.

Based on these assumptions, we derived the following hypotheses:

1. For both emotion and gender evaluation, we can identify networks that are driven by attention: specific for the task (emotion/gender), but independent from stimulus modality or memory representations.
2. For both emotion and gender evaluation, we can identify networks that are driven by supramodal memory representations: independent from stimulus modality, but only present for associated auditory and visual stimuli.

These hypotheses were tested using fMRI.

2 | MATERIALS AND METHODS

2.1 | Participants

Forty-five right-handed subjects (23 female; age span 19–33 years, mean 24.7 ± 3.1) participated in the experiment. All subjects had normal or corrected to normal vision, normal hearing, no contraindications against MR investigations, and no history of neurological or psychiatric illness. All participants had either German as a first language or were grown up bilingually (with German from early childhood on).

The experiment was designed according to the Code of Ethics of the World Medical Association (Declaration of Helsinki, 2013, and the study protocol was approved by the Ethics Committee of the Medical Faculty at RWTH Aachen University (EK 003/14). After complete description of the study to the subjects, written informed consent was obtained.

2.2 | Stimuli

Auditory stimuli were disyllabic pseudowords (Thonnessen et al., 2010). They followed German phonological rules but had no semantic content and were validated in a pre-study on 25 subjects who did not participate in the fMRI study (see Klasen et al., 2011 for details of stimulus validation). Auditory stimulus duration was 1 s. Visual stimuli were taken from the validated NimStim Face Stimulus Set (Tottenham et al., 2009). In analogy to the duration of auditory stimuli, photographs were presented for 1 s each. Stimuli were always presented in isolation (unimodal presentation). Auditory and visual stimuli were counterbalanced for emotion (50% happy, 50% angry) and gender of the speaker/actor (50% female, 50% male). Moreover, each stimulus type was displayed by four different speakers/actors. In summary, the experiment thus comprised 32 different stimuli: 2 modalities (auditory/visual) \times 2 emotions (happy/angry) \times 2 genders (female/male) \times 4 actors/speakers.

2.3 | Experimental design

In the present fMRI study, we employed a SIAT. The SIAT measures crossmodal associations between corresponding visual and auditory modalities of social signals (faces and voices) via reaction times. Similar to the original IAT, the SIAT uses a classification task with two target categories sharing one response key, paired in either congruent or incongruent fashion. In the congruent condition, corresponding visual and auditory signals (e.g., happy faces and happy voices) share the same response key, whereas in the incongruent condition non-matching pairings (e.g., happy faces and angry voices) are mapped on the same key.

To address the top-down influence of attentional focus on social perception, two different SIAT variants were employed: an emotion SIAT, and a gender SIAT. Attentional focus was varied via the task. In the Emotion SIAT, the task was to classify the emotion of faces and voices (happy or angry), and in the Gender SIAT, the task was to classify stimulus gender (male or female). The task of the participant was to classify the stimuli according to the respective instruction as fast and as accurately as possible by pressing one of two response keys according to the assigned category. Pairings of target categories were either congruent or incongruent. In the congruent condition, corresponding auditory and visual stimuli were always mapped on the same key, for example, for emotion, angry voice and angry face on one key and happy face and happy voice on the other key. In the incongruent condition, non-corresponding auditory and visual stimuli

were mapped on the same key, for example, for emotion angry face and happy voice on one key and happy face and angry voice on the other key. The gender SIAT was designed in analogy.

Both SIATs included one congruent and one incongruent association phase in separate sessions in randomized order. Prior to the first association phase, participants performed two shorter learning phases, where the assignment of keys for the categories was learned according to the first association phase. A learning phase consisted of either visual or auditory stimuli only. As an example, a congruent association phase was always preceded by two learning phases assigning auditory and visual emotions in a congruent fashion (e.g., happy voice = left, angry voice = right for the auditory learning phase and happy face = left and angry face = right for the visual learning phase in the emotion SIAT). The two association phases were separated by an additional re-learning phase (either auditory or visual) with the identical setup, but with switched assignments of keys, preparing for the second association phase (see Figure 1 for a depiction of the experimental setup).

The order of the SIAT variants (emotion/gender) was randomized for each participant. The same was true for the order of the association phases within each SIAT (congruent/incongruent), the order of the learning phases (visual/auditory), and the assignment of emotion (angry/happy) and gender (male/female) to the response keys (right/left).

In the SIAT, reaction time differences between the congruent and incongruent association tasks quantified the implicit association of auditory and visual representations of the social categories emotion and gender. Both SIATs were conducted in a repeated measurement

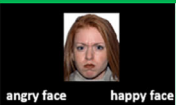
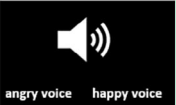



		Original IAT (Greenwald et al., 1998)	Social IAT 2 variants: Emotion evaluation / Gender evaluation				
			Parallel versions				
Phase I	Learning A	Social group evaluation	Auditory stimulus evaluation	Visual stimulus evaluation	Auditory stimulus evaluation	Visual stimulus evaluation	
Phase II	Learning B	Attribute evaluation	Visual stimulus evaluation	Auditory stimulus evaluation	Visual stimulus evaluation	Auditory stimulus evaluation	
Phase III	Association A	Combined task A	Combined task A	Combined task A	Combined task A	Combined task A	
Phase IV	Re-Learning	Attribute evaluation	Auditory stimulus evaluation	Visual stimulus evaluation	Visual stimulus evaluation	Auditory stimulus evaluation	
Phase V	Association B	Combined task B	Combined task B	Combined task B	Combined task B	Combined task B	

FIGURE 1 Experimental design. Two Social Implicit Association Test (SIAT) variants were employed: one for emotion evaluation and one for gender evaluation. Both SIATs consisted of five phases, in close analogy to the original IAT by Greenwald et al. (1998). To avoid sequential effects of auditory and visual evaluation phases, four parallel versions were employed for each of the SIATs (emotion/gender). Inserts on the right show one example version of the emotion evaluation SIAT in detail

design on two different days. Auditory and visual stimuli were identical in both SIATs.

Although the original IAT has traditionally been used to measure attitudes (stereotypes/implicit bias) in social psychology (e.g., Gawronski, 2002; Wilson & Scior, 2013), research has shown that adaptations of the IAT paradigm can be used for associations between non-social categories as well (e.g., flowers/insects and their association with pleasant/unpleasant attributes; Greenwald et al., 1998). Moreover, the IAT works for the auditory domain as well (McKay, Arciuli, Atkinson, Bennett, & Pheils, 2010) and even for the association between auditory and visual domains (Parise & Spence, 2012). This universal applicability encouraged us to use the SIAT as a social variant for investigating associations between vocal and facial stimuli.

In summary, the SIAT design allowed us to investigate the top-down contributions of attentional focus and memory representation independently from each other. By using conjunction analyses, we were moreover able to identify activation patterns that were independent from stimulus modality (i.e., supramodal). To avoid any bias arising from the stimulus material itself (and thus to exclude any bottom-up effects), we used exactly the same stimulus material for all association tasks.

Images were presented through a mirror mounted on the head coil. During the fMRI measurements, participants wore soft foam ear plugs and head phones, which served as ear protection as well as for delivering the auditory stimuli. The sound volume was tested before the measurements in the scanner and individually adjusted to a comfortable level, based on the participant's feedback. Previous experience with the same scanner, ear protection, and auditory stimulus set (e.g., Klaser et al., 2011) indicated that the stimuli were well audible and could easily be classified even with the scanner noise in the background. Responses were given via two keys on a keypad placed at the participant's right hand.

2.4 | Data acquisition

Whole-brain fMRI was conducted with echo-planar imaging (EPI) sequences (TE = 28 ms, TR = 2,000 ms, flip angle = 77°, voxel size = 3 × 3 mm, matrix size = 64 × 64, 34 transverse slices, 3 mm slice thickness, 0.75 mm gap) on a 3 Tesla Siemens Prisma MRI scanner (Siemens Medical, Erlangen, Germany) using a 12-channel head coil. The learning phases comprised 110 volumes each; association phases comprised 390 volumes. After the functional measurements, high-resolution T1-weighted anatomical images were performed using a magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (TE = 2.52 ms; TR = 1,900 ms; TI = 900 ms; flip angle = 9°; FOV = 256 × 256 mm²; 1 mm isotropic voxels; 176 sagittal slices). Total time for functional and anatomical scans was 45 min.

2.5 | Data analysis

Image analyses were performed with BrainVoyager QX 2.8 (Brain Innovation, Maastricht, The Netherlands). Preprocessing of the functional MR images included slice time correction, 3D motion correction,

Gaussian spatial smoothing (6 mm full width half maximum kernel), and high-pass filtering including linear trend removal. The first five images of each functional run were discarded to avoid T1 saturation effects. Functional images were coregistered to 3D anatomical data and transformed into Talairach space (Talairach & Tournoux, 1988), following the standard procedure as implemented in BrainVoyager. In total, four participants were excluded from the analysis. One was excluded due to technical problems; a part of the original DICOM image files was damaged and could not be restored. Three additional participants were excluded from all further analyses due to excessive head motion as identified by visual inspection, leaving a total of 41 participants in the final sample. From the excluded participants, two were male and two were female, leaving a final sample of 21 female and 20 male participants.

Statistical parametric maps were created by using a random effects general linear model (RFX-GLM) with multiple predictors according to the stimulus types. The following within-subject factors were considered in the analysis:

Attentional focus (**Emotion** vs. **Gender**)

Congruency of target categories (**Congruent** vs. **Incongruent**)

Stimulus modality (**Face** vs. **Voice**)

Stimulus Gender (**Female** vs. **Male**)

Stimulus Emotion (**Happy** vs. **Angry**)

The full combination of these five factors led to a total of $2^5 = 32$ predictors which are listed in Table 1 (abbreviations see above). For each of the contrasts, their encoding is marked with "+" and "-", respectively.

Fixation cross phases served as a low-level baseline. Events were defined in a stimulus-bound fashion, that is, modeled for the duration of stimulus presentation. Only trials with correct responses were included in the analyses. Trials with missing or incorrect responses were modeled as separate confound predictors. Task contrasts were investigated via paired t tests. Following the recommendations of Woo, Krishnan, and Wager (2014), activations were thresholded at voxel-wise $p < .001$ and Monte-Carlo-corrected for multiple comparisons on the cluster level ($p < .05$, corresponding to $k > 11$). All reported conjunction analyses tested the conservative conjunction null hypotheses (Nichols, Brett, Andersson, Wager, & Poline, 2005). Results are displayed in radiological convention (left is right).

To address the study's hypotheses, the following comparisons were of interest:

1. *Emotion versus gender: Supramodal networks.* These were networks specific for emotion resp. gender evaluation, but independent from stimulus modality. These were the contrasts (Voice Emotion > Voice Gender) \cap (Face Emotion > Face Gender) as well as the reversed contrast (Gender > Emotion, respectively).
2. *Emotion versus gender: Networks independent from congruency of target categories.* These were networks specific for emotion resp. gender evaluation, but independent from congruency of the target categories. These were the contrasts (Congruent Emotion > Congruent Gender) \cap (Incongruent Emotion > Incongruent Gender) as well as the reversed contrast (Gender > Emotion, respectively).

TABLE 1 Predictors and their encoding

Predictor name	Figure 2a Blue	Figure 2a Red	Figure 2b Green	Figure 2b Yellow	Figure 4 Emotion	Figure 4 Gender	Figure 5 Blue	Figure 5 Red
EmCoFaFeHa		+	+		+			
EmCoFaMaHa		+	+		+			
EmCoFaFeAn		+	+		+			
EmCoFaMaAn		+	+		+			
EmCoVoFeHa	+		+		+			
EmCoVoMaHa	+		+		+			
EmCoVoFeAn	+		+		+			
EmCoVoMaAn	+		+		+			
EmlnFaFeHa		+		+		-		
EmlnFaMaHa		+		+		-		
EmlnFaFeAn		+		+		-		
EmlnFaMaAn		+		+		-		
EmlnVoFeHa	+			+		-		
EmlnVoMaHa	+			+		-		
EmlnVoFeAn	+			+		-		
EmlnVoMaAn	+			+		-		
GeCoFaFeHa		-	-			+		+
GeCoFaMaHa		-	-			+		+
GeCoFaFeAn		-	-			+		+
GeCoFaMaAn		-	-			+		+
GeCoVoFeHa	-		-			+	+	
GeCoVoMaHa	-		-			+	+	
GeCoVoFeAn	-		-			+	+	
GeCoVoMaAn	-		-			+	+	
GelnFaFeHa		-		-		-		-
GelnFaMaHa		-		-		-		-
GelnFaFeAn		-		-		-		-
GelnFaMaAn		-		-		-		-
GelnVoFeHa	-			-		-	-	
GelnVoMaHa	-			-		-	-	
GelnVoFeAn	-			-		-	-	
GelnVoMaAn	-			-		-	-	

3. *Emotion versus gender: Networks depending exclusively on attentional focus.* These were networks depending solely on the attention focus (emotion or gender), independent from stimulus modality or congruency of the target concepts. This equals to the fourfold conjunction (Voice Emotion > Voice Gender) \cap (Face Emotion > Face Gender) \cap (Congruent Emotion > Congruent Gender) \cap (Incongruent Emotion > Incongruent Gender) as well as the reversed contrast (Gender > Emotion, respectively).
4. *Networks depending on the congruency of target categories.* These were networks that were specific for congruency resp. incongruency of the target categories (congruent vs. incongruent and vice versa). They were investigated separately for emotion and gender evaluation, as well as for the comparison between them.

5. *Networks of supramodal memory representations.* Networks independent from stimulus modality, but depending on the congruency of the target concepts. This corresponds to the conjunction (Voice Congruent > Voice Incongruent) \cap (Face Congruent > Face Incongruent), calculated for both emotion and gender classification.

3 | RESULTS

3.1 | Behavioral results

For emotion evaluation, 92.38% (118.24 \pm 5.82) of all stimuli were classified correctly when target concepts were congruent, whereas

90.15% (115.39 ± 6.34) were classified correctly when target concepts were incongruent. Average reaction times for emotion evaluation were 985.90 ± 145.55 ms (congruent) and $1,150.01 \pm 186.80$ ms (incongruent). For gender evaluation, 95.26% (121.93 ± 5.38) of all stimuli were classified correctly when target concepts were congruent, whereas 94.40% (120.83 ± 4.92) were classified correctly when target concepts were incongruent. Average reaction times for gender evaluation were 859.63 ± 134.28 ms (congruent) and $1,012.68 \pm 159.25$ ms (incongruent).

In the SIAT—as well as in the original IAT—the Implicit Association Effect is quantified via reaction time differences between different pairings of target concepts. Increased reaction times reflect weaker associations between the investigated concepts and thus higher task difficulty. To investigate effects of target congruency, evaluation task, and their interaction on reaction times, we therefore calculated a 2×2 ANOVA with the factors Task (emotion evaluation/gender evaluation) and Congruency of target concepts (congruent/incongruent). Results revealed significant main effects of Task ($F[1, 40] = 69.35$, $p < .001$) and Congruency of target concepts ($F[1, 40] = 174.96$, $p < .001$), but no interaction ($F[1, 40] = 0.40$, $p = .51$).

RTs for the four association conditions (Emotion evaluation congruent, Emotion evaluation incongruent, Gender evaluation congruent,

Gender evaluation incongruent) were tested for normal distribution. In all four conditions, reaction times were normally distributed (Kolmogorov–Smirnov test, $p(\text{EmoCong}) = .20$, $p(\text{EmoInkong}) = .14$, $p(\text{GendCong}) = .20$, $p(\text{GendInkong}) = .16$). Further data trimming was not performed. RTs and their *SD* in the present study were highly congruent with the data from a pre-study on 42 subjects with very similar demographics without fMRI, using the exactly same paradigm. The comparison of the two data sets thus suggests a high level of replicability of the results.

3.2 | Neuroimaging results

3.2.1 | Emotion versus gender: Supramodal networks

These were networks specific for emotion resp. gender evaluation, but independent from stimulus modality. We thus compared emotion versus gender evaluation networks separately for auditory (voice) and visual (face) stimuli.

Emotion > gender. For emotion evaluation, auditory stimuli involved inferior frontal areas, along with pre-supplementary motor area (pre-SMA) and anterior insula (Figure 2a; blue). Visual stimuli

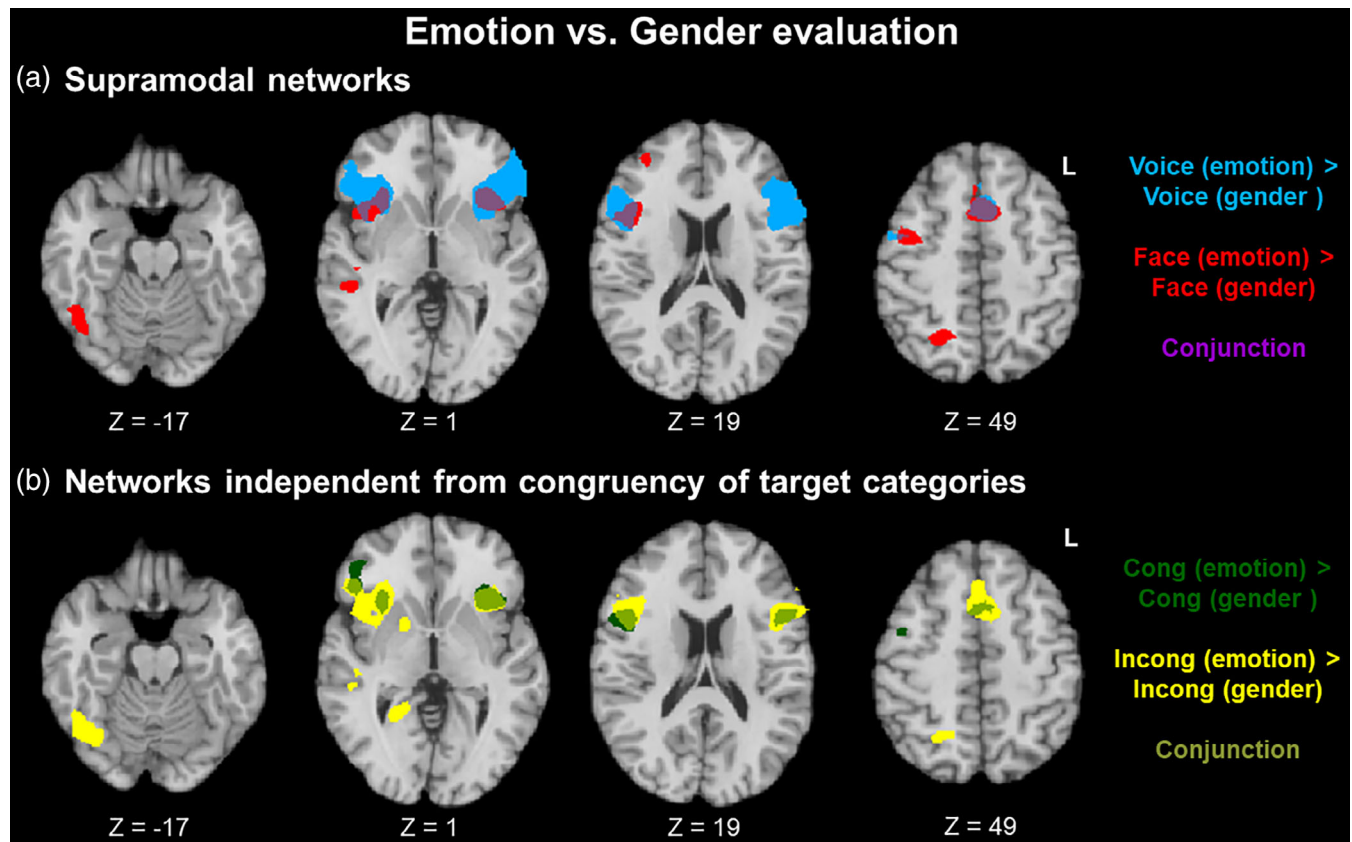


FIGURE 2 Emotion versus gender evaluation. (a) Supramodal networks. Besides some modality-specific patterns such as fusiform face area for facial stimuli (red), emotion evaluation—as compared to gender evaluation—invoked a modality-independent network in bilateral anterior insula, right inferior frontal gyrus (IFG), and pre-supplementary motor area (SMA) (a; purple). A similar emotion evaluation network emerged independently from the congruency of the target categories (b; light green)

enhanced activation in right fusiform face area (FFA) and superior temporal sulcus (STS), along with right inferior frontal gyrus (IFG) and middle frontal gyrus (MFG), pre-SMA, and anterior insula (Figure 2a; red). A conjunction of both maps was observed in bilateral anterior insula, right IFG, and pre-SMA (Figure 2a; purple). These areas thus reflect networks for emotion evaluation that are independent from stimulus modality.

Gender > emotion. For gender evaluation, auditory stimuli involved VMPFC and ACC, along with left angular and superior frontal gyri (Figure S1). No clusters emerged for visual stimuli or for the conjunction of both contrasts.

3.2.2 | Emotion versus gender: Networks independent from congruency of target categories

These were networks specific for emotion resp. gender evaluation, but independent from congruency of the target categories.

Emotion > gender. For emotion evaluation, the congruent condition involved bilateral IFG, pre-SMA, and bilateral anterior insula (Figure 2b; dark green), whereas the incongruent condition enhanced activation in globus pallidus, right FFA and STS, bilateral IFG, pre-SMA, and bilateral anterior insula (Figure 2b; yellow). A conjunction of both maps was observed in bilateral anterior insula, right IFG, and pre-SMA (Figure 2b; light green). These areas thus reflect networks for emotion evaluation that are independent from the congruency of target concepts.

Gender > emotion. For gender evaluation, congruent target categories involved left angular gyrus (Figure S2). No clusters emerged for incongruent target categories or for the conjunction of both contrasts.

3.2.3 | Emotion versus gender: Networks depending exclusively on attentional focus

These were networks depending solely on the attention focus (emotion or gender), independent from stimulus modality or congruency of the target concepts.

Emotion > gender. To investigate effects specific for emotion evaluation independently from stimulus modality and from the congruency of target concepts, we thus calculated the fourfold conjunction of all maps, that is, (Voice emotion > Voice gender) \cap (Face emotion > Face gender) \cap (Congruent emotion > Congruent gender) \cap (Incongruent emotion > Incongruent gender). The resulting map revealed a common activation in bilateral anterior insula, right IFG, and pre-SMA (Figure 3; Table 1).

Differences in reaction times between emotion and gender evaluation indicated a higher difficulty of the emotion task. To investigate possible influences of the latter on the brain activation patterns displayed in Figure 3, we calculated a subject-wise task difficulty coefficient for this contrast, which was defined as (AverageRTEmotion – AverageRTGender). This coefficient was correlated with individual contrast values in all clusters. Even at a liberal threshold of $p < .05$ and without correction for multiple comparisons, no correlation with task difficulty was observed in any of the clusters (IFG R: $r(39) = .01, p = .94$; Ant Ins R: $r(39) = -.16, p = .31$; SMA: $r(39) = -.25, p = .12$; Ant Ins L: $r(39) = -.04, p = .80$). For the regions of this network, we moreover compared contrast values between different stimulus types with paired t tests (Figure 3). Notably, angry voices constantly yielded the strongest effect in all four regions, whereas no difference was observed between any of the other three stimulus types (happy voice, angry face, happy face).

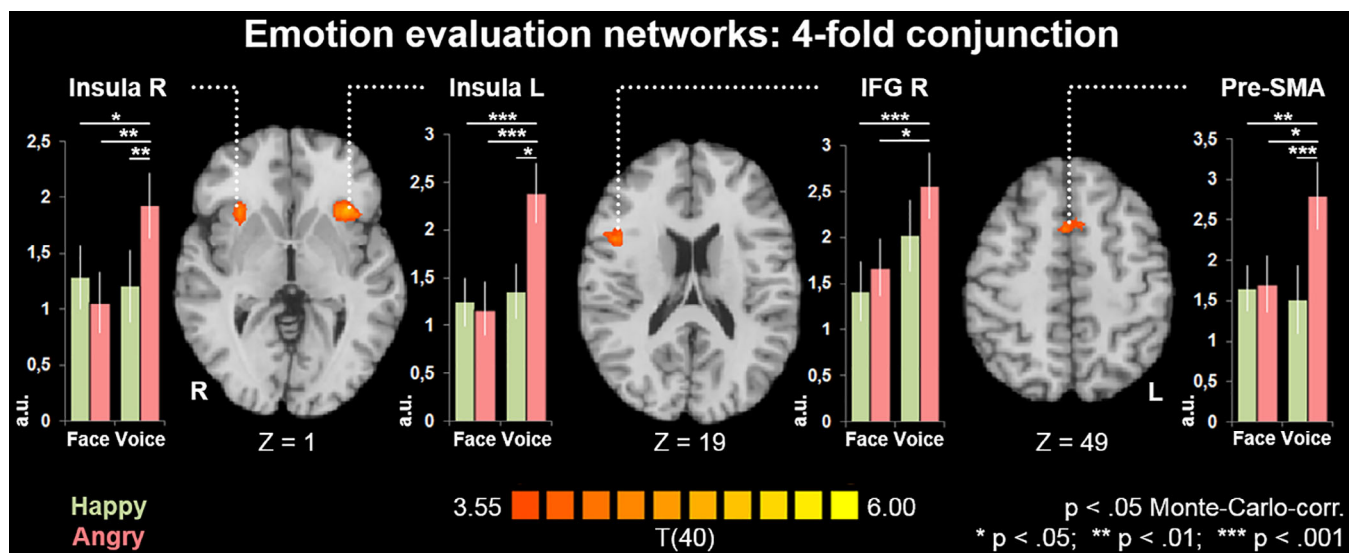


FIGURE 3 Emotion evaluation networks: fourfold conjunction. The fourfold conjunction (Voice emotion > Voice gender) \cap (Face emotion > Face gender) \cap (Congruent emotion > Congruent gender) \cap (Incongruent emotion > Incongruent gender) confirmed bilateral anterior insula, right inferior frontal gyrus (IFG), and pre-supplementary motor area (SMA) as an emotion evaluation network irrespective of stimulus modality and congruency of target concepts. This network thus reflects the top-down contribution of attentional focus during emotion evaluation. In all regions of this network, angry voices evoked the strongest response (a.u. = arbitrary units, derived from mean beta values from the first level general linear models [GLMs]). No such attention-driven network was observed for gender evaluation

Gender > emotion. No clusters emerged for the fourfold conjunction (Voice gender > Voice emotion) \cap (Face gender > Face emotion) \cap (Congruent gender > Congruent emotion) \cap (Incongruent gender > Incongruent emotion).

3.2.4 | Networks depending on the congruency of target categories

These were networks that were specific for congruency resp. incongruency of the target categories (congruent vs. incongruent and vice versa).

Emotion evaluation. Incongruence led to a stronger activation in areas of the emotion evaluation network (compare Figure 2), namely FFA, bilateral anterior insula, thalamus, globus pallidus, IFG/MFG, and pre-SMA, along with extended activation in visual systems, DLPFC,

and superior parietal lobule (SPL; Figure 4). Congruency, in contrast, was not associated with any specific activation pattern during emotion evaluation.

Gender evaluation. A different picture emerged for the gender evaluation task. The incongruent condition led to a similar, albeit less pronounced pattern in pre-SMA, DLPFC, right anterior insula, and SPL. Congruency, in turn, led to enhanced activation in two prominent clusters in rACC and VMPFC (Figure 4).

3.2.5 | Networks of supramodal memory representations

These were networks independent from stimulus modality, but depending on the congruency of the target concepts.

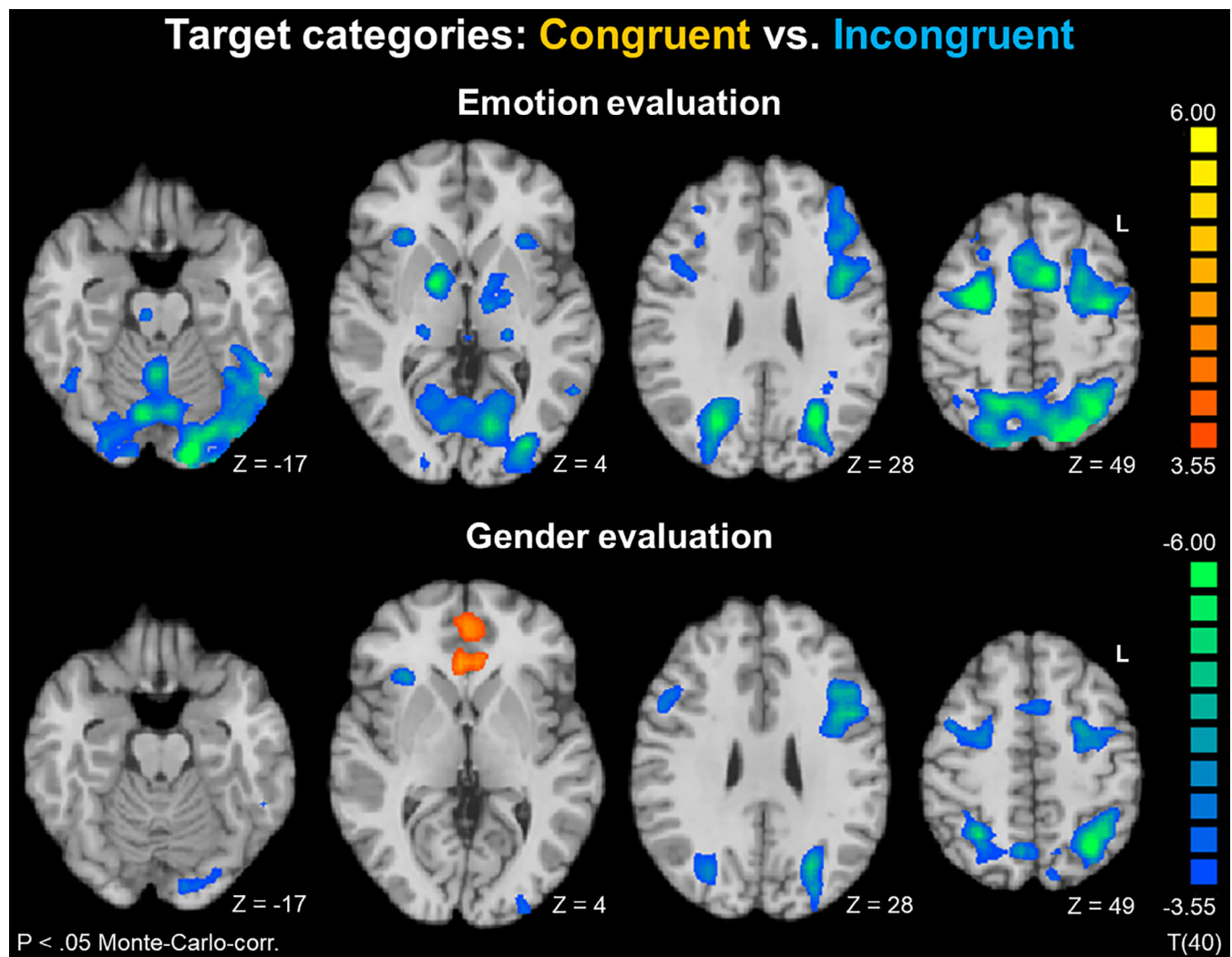


FIGURE 4 Target categories: Congruent versus incongruent. Both emotion and gender evaluation showed similar fronto-parietal networks for incongruent target categories, indicating increased working memory load. No congruency-specific activation was observed for emotion evaluation. For gender evaluation, congruency led to enhanced activation in rostral anterior cingulate cortex (rACC) and ventromedial prefrontal cortex (VMPFC). These networks may thus reflect supramodal memory representations supporting gender evaluation

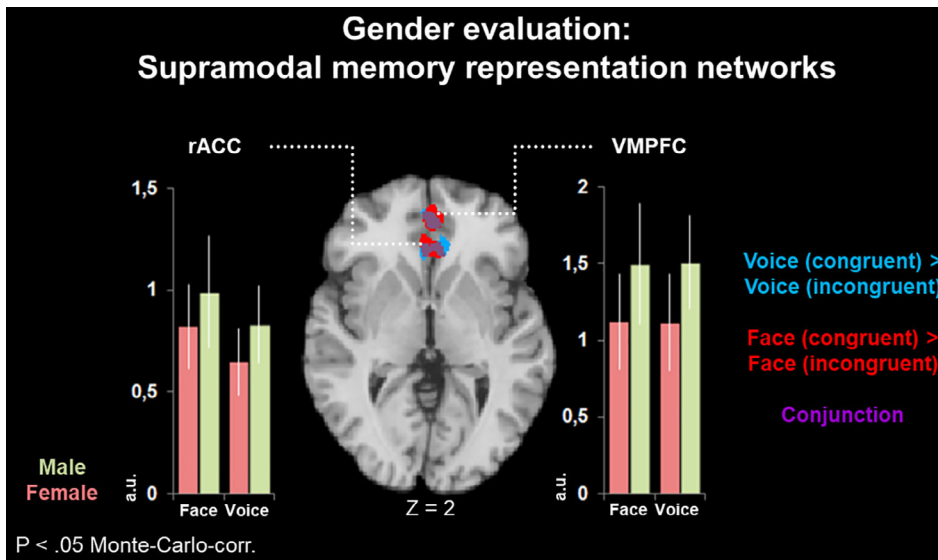


FIGURE 5 Gender evaluation: Supramodal memory representation networks. For gender evaluation, congruency-specific networks in rostral anterior cingulate cortex (rACC) and ventromedial prefrontal cortex (VMPFC) were independent from stimulus modality (conjunction of Voice and Face; purple), thus confirming the contribution of a supramodal memory network. Within this network, no significant differences emerged between male and female stimuli (a.u. = arbitrary units, derived from mean beta values from the first level general linear models [GLMs]). For emotion evaluation, no such network was observed

Emotion evaluation. No clusters emerged for the conjunction (Voice congruent > Voice incongruent) \cap (Face congruent > Face incongruent).

Gender evaluation. The maps for auditory (blue) and visual stimuli (red) largely overlapped in both clusters (Figure 5), as reflected by the conjunction (Voice congruent > Voice incongruent) \cap (Face congruent > Face incongruent), displayed in purple. In both regions, contrast values were compared with paired t tests. Within each region, no differences emerged between different evaluated stimulus types (male and female faces and voices). The clusters from the functional maps in Figures 2-5 are listed in Table 2.

4 | DISCUSSION

The present study revealed new insights into top-down contributions to social perception. Specifically, the SIAT enabled us to identify top-down influences in the processing of social information in the auditory and visual domains. Task-specific, but modality-independent patterns reflected supramodal networks for social categories. These top-down components could further be separated into attention-driven networks and supramodal memory representations. Functionally distinct networks were identified for the social categories emotion and gender.

For emotion evaluation, the modality-specific analysis revealed FFA and STS specifically for face processing. Besides the FFA's well-established role in face identification, which is assumed to rely on mainly invariant facial features (Calder & Young, 2005; Dekowska, Kunięcki, & Jaśkowski, 2008; Haxby, Hoffman, & Gobbini, 2002; Hoffman & Haxby, 2000), recent studies highlight the importance of the FFA for processing emotional expressions as well (Harry, Williams, Davis, & Kim, 2013; Nestor, Plaut, & Behrmann, 2011). In line with these findings, our study supports the notion of the FFA as part of a visual emotion recognition network. Moreover, we extend present findings by demonstrating a task dependency of the stimulus-evoked

activation. Specifically, FFA activity to facial stimuli was observed only during emotion, but not during gender classification, therefore reflecting top-down modulation. Evaluative networks in the FFA thus seem to rely on the attentional focus.

Moreover, we confirmed the role of the posterior right STS for conscious processing of visual emotions. The STS was traditionally regarded as a flexible component in different neural pathways without a specialized functional attribution (for a review see Hein & Knight, 2008). Indisputably, the STS has various functions including Theory of Mind abilities, social perception, and multisensory integration (Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005; Campanella & Belin, 2007; Saxe, 2006). This functional versatility may be explained by sub-regional specialization, but also by task-dependent co-activation with functionally distinct frontal and temporal networks (Hein & Knight, 2008). Functional synchronicity of posterior STS with the FFA may reflect a visual emotion processing network.

Current models of auditory emotion processing highlight a right-hemispheric lateralization (Brück, Kreifelts, & Wildgruber, 2011; Klasen et al., 2018). Right sided primary and higher order acoustic regions extract suprasegmental information, followed by processing of meaningful suprasegmental sequences in posterior parts of the right STS, followed by evaluation of emotional prosody in IFG (Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Neuroimaging findings (Klasen et al., 2018) highlight the relevance of right IFG for emotional prosody. In our study, a right-hemispheric lateralization was observed for the fourfold conjunction of all maps (Figure 3c), showing specific effects of emotion evaluation independently from modality and congruency of target concepts. Extending previous findings, our study highlighted the role of right IFG in emotion recognition from both auditory and visual domains. From an integrative perspective, there may thus be a supramodal right hemispheric dominance for emotion processing (cf. Le Grand, Mondloch, Maurer, & Brent, 2003).

Functional mapping of explicit emotion processing furthermore revealed effects in bilateral anterior insula and pre-SMA. Research has

TABLE 2 Clusters from mapping in Figures 2–5

Cluster	Brain region	TAL coordinates			Peak T	mm ³
		x	y	z		
Figure 2a: conjunction voice and face						
1	Inferior frontal gyrus r	42	20	16	4.95	3,048
2	Insula r	30	17	1	4.61	1,563
3	Pre-supplementary motor area r/l	−6	11	49	5.49	2,051
4	Insula l	−33	20	1	5.28	1,961
Figure 2b: Conjunction congruent and incongruent						
1	Inferior frontal gyrus r	45	11	16	4.56	1,738
2	Insula r	30	20	1	4.78	544
3	Pre-supplementary motor area r/l	6	11	61	4.53	821
4	Insula l	−33	20	1	5.50	1,978
5	Inferior frontal gyrus l	−45	14	19	5.26	630
Figure 3: Fourfold conjunction						
1	Inferior frontal gyrus r	45	11	16	4.56	1,010
2	Insula r	30	17	1	4.45	510
3	Pre-supplementary motor area r/l	6	11	61	4.53	808
4	Insula l	−33	20	1	5.28	1,564
Figure 4: Congruent > incongruent target categories						
Emotion						
1	Inferior frontal gyrus r, Insula r	30	23	7	−5.32	4,165
2	Fusiform gyrus r	45	−58	17	−4.19	297
3	Inferior parietal lobule r/l, superior parietal lobule l/r, Cuneus r/l, lingual gyrus r/l, inferior/middle occipital gyrus r/l, fusiform gyrus l, brainstem r/l, thalamus r/l, Cerebellum r/l	27	−61	−17	−8.74	133,612
4	Precentral gyrus r, superior frontal gyrus r, middle frontal gyrus r	27	−4	49	−7.89	8,409
5	Superior frontal gyrus r, middle frontal gyrus r	27	23	43	−4.66	2,384
6	Thalamus r, Globus pallidus r	15	−4	1	−6.76	4,717
7	Inferior frontal gyrus l, Insula l, precentral gyrus r, superior frontal gyrus r, middle frontal gyrus r, pre-supplementary motor area r/l	−33	−4	58	−7.20	32,879
8	Thalamus r, Globus pallidus r	−21	−22	−2	−5.23	3,392
Gender						
1	Precentral gyrus r, superior frontal gyrus r, middle frontal gyrus r	36	−4	43	−5.07	5,084
2	Insula r	36	−13	19	−4.33	441
3	Insula r	27	23	7	−4.94	535
4	Superior parietal lobule r/l, inferior parietal lobule l/r, Precuneus l/r	−30	−55	43	−8.11	28,037
5	Cerebellum r/l	9	−70	−26	−4.87	1,251
6	Precentral gyrus l, superior frontal gyrus l, middle frontal gyrus l	−42	2	37	−7.57	12,501
7	Ventromedial prefrontal cortex r/l	0	47	4	4.76	2,735
8	Rostral anterior cingulate gyrus r/l,	0	32	−2	4.99	2,405
9	Cuneus l, lingual gyrus l	−12	−88	−14	−5.00	3,570
10	Fusiform gyrus l	−45	−52	−11	−4.48	836
Figure 5: Conjunction voice and face						
1	Ventromedial prefrontal cortex r/l	0	47	1	3.98	736
2	Rostral anterior cingulate cortex r/l	−3	32	−2	4.27	566

shown that the pre-SMA is involved in domain-general sequence processes (Cona & Semenza, 2017) and in emotional evaluation of signals irrespective of modality (Ethofer et al., 2013). The anterior insula has been ascribed to a wide range of complex functions and participates in various cognitive and emotional processes (for a review see Menon & Uddin, 2010). In line with our findings, Menon and Uddin (2010) propose that a basic function of the anterior insula is the bottom-up driven detection of salient stimuli across multiple modalities. It is well established that the insula engages in affective processes (e.g., emotion perception of others) and the experience of emotions that derive from visceral and somatic information about bodily states (Uddin, 2015). As such, insula activity represents an individual's subjective and conscious emotional state, as well as the emotive value of external stimuli. Thus, it has been suggested that the ability to understand the emotions of others depends largely on experiencing similar changes in our visceral state by mirroring the perceived emotion (Critchley & Harrison, 2013). The anterior insula may be a central hub in this function. Taken together, the observed activity of SMA and anterior Insula may represent a supramodal neuronal signature of explicit emotion processing. However, since similar patterns emerged for the congruent as well as for the incongruent target concept they reflect an evaluative rather than a supramodal memory network. Considering the similarity with the salience network (Menon, 2015), the pattern reflects the high evolutionary significance of emotion recognition.

Notably, angry voices elicited the strongest responses in the emotion evaluation network. Previous research revealed an overall increase in activation for vocal emotion compared with neutral expressions in a fronto-temporo-striatal network (Ethofer et al., 2006; Kotz et al., 2003). Ethofer et al. (2009) investigated brain regions that were more responsive to angry than to neutral prosody and identified bilaterally IFG/OFC, amygdala, insula, mediodorsal thalamus, and the middle part of the STG. Furthermore, they showed that the activation of these regions was automatic and independent of the underlying task, concluding that angry prosody is processed irrespectively of cognitive demands and attentional focus. Similar findings can be observed for visual emotion processing. Vuilleumier (2005) inferred that the FFA was more activated by fearful than neutral faces, even when faces were task-irrelevant. Our findings support the notion that angry prosody is perceived with particular dominance, which is of fundamental importance to prioritize the procession of threat-related stimuli (Cox & Harrison, 2008; LeDoux, 2003).

Remarkably, no amygdala activation was observed for any of the emotion classification categories. Lesion studies show that the amygdala has modulatory influences on emotion processing areas and heightens activity in for example, the FFA when perceiving fearful faces compared to neutral (Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004), and this is also true for prosodic emotion processing and the STS (Frühholz et al., 2015). Since emotional information was present in all trials, missing amygdala differences can be attributed to unattentive emotion processing in all tasks. This notion is supported by previous findings (Vuilleumier, 2005; Vuilleumier, Armony, Driver, & Dolan, 2001) and was also explicitly validated in trials with a gender

classification task, where amygdala activation was present even though attention was directed to the gender (Morris, Ohman, & Dolan, 1998).

A comparison of congruent versus incongruent target concepts revealed increased workload in a fronto-parietal network for incongruence in emotion and gender SIATs. This network has already been described for the evaluation of semantic incongruent bimodal emotional stimuli (Klasen et al., 2011). It shows a large overlap with the executive control network as described by Seeley et al. (2007), which reflects attention, working memory, and response selection. Almost identical findings between our study and Klasen et al. (2011) indicate a negligible influence of stimulus modality. Instead, the network seems to be driven by the aspect of incongruence itself, putatively reflecting increased task difficulty and cognitive workload in the incongruent condition. Moreover, pre-SMA activity may reflect conflict monitoring and error detection (Mayer et al., 2012). In a similar way, social classification categories (emotion vs. gender) seem to be only of minor importance for incongruence networks.

In contrast to our initial hypotheses, we could not reveal a contribution of a supramodal memory representation for emotional categorization. Instead, emotion evaluation seems to involve large evaluative networks, some of them modality-independent, others not. In summary, recognition of facial and vocal emotions involves common networks in insula, IFG, and pre-SMA, but does not rely on a common supramodal memory representation. In line with this notion, a recent meta-analysis by Schirmer (2018) revealed fundamentally different pathways for auditory and visual emotions. Effects of effortful, that is, conscious emotion processing were observed in supplementary motor regions, which is in line with our findings. Emotion processing effects in limbic areas such as the amygdala, in turn, were task-independent and largely driven by the visual modality (Schirmer, 2018). This also delivers a new perspective on crossmodal emotion integration. Neuroimaging findings show that congruent audiovisual emotions enhance activity primarily in limbic areas (Klasen et al., 2011). In line with the well-established visual dominance effect in audiovisual perception (Colavita, 1974), auditory emotions may be just a supplement to visual perception, both behaviorally and neurobiologically, without the need for recruiting a common memory representation.

The gender SIAT, in contrast, showed enhanced involvement of two prominent clusters: the VMPFC and ACC. These findings are well in line with the findings from Knutson et al. (2007) on the neural substrates of gender and racial bias, as well as with Milne and Grafman (2001), who found a reduced IAT effect in patients with VMPFC lesions. In these studies, VMPFC was considered as representing previously learned automatic processing of emotional and social information. Thus, VMPFC may support concept formation in long-term memory.

Widening the scope, this view is very much in line with neuroimaging research on schematic memory. Schemas are experience-based implicit memory representations of situational aspects that typically belong together. They are activated by perceptual input and form a framework for stimulus interpretation (Bowman & Zeithamova, 2018;

Spalding, Jones, Duff, Tranel, & Warren, 2015), a conceptualization closely related to the spreading activation network theory by Collins and Loftus (1975). Recent neuroimaging studies highlight VMPFC contributions to establishing and retrieving schemas. A lesion study by Spalding et al. (2015) indicated reduced performance ability of subjects with focal VMPFC damage for integrating new information into a schema congruent context. In a recent fMRI study, Bowman and Zeithamova (2018) describe the VMPFC as representing abstract prototype information, supporting generalization in conceptual learning over multiple domains. In summary, the VMPFC seems to store memories about typical examples and characteristic features of object categories. These “prototype” representations seem to facilitate object recognition in a top-down fashion: classification and response selection are based on the comparison of perceptual input with memory prototypes. In the case of gender, facial and vocal stimuli seem to access the same supramodal memory prototype, which may also account for the enhanced accuracy and reaction times compared to the emotion classification task. Supramodal prototypes may exist for emotions as well; however, the present study found no evidence for their contribution to stimulus classification.

5 | CONCLUSION

The present study identified modality-specific and modality-independent influences of attentional focus and memory representations on the neural processing of social stimuli. Irrespective of modality, emotion evaluation engaged a fronto-insular network which was independent from supramodal memory representations. Gender classification, in turn, relied on supramodal memory representations in rACC and VMPFC.

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft [DFG], IRTG 1328 and IRTG 2150), the Federal Ministry of Education and Research (APIC: 01EE1405B), and the Brain Imaging Facility of the Interdisciplinary Center for Clinical Research (ICCR) Aachen. Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham at tott0006@tc.umn.edu for more information concerning the stimulus set.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from RWTH Aachen University. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the corresponding author with the permission of RWTH Aachen University.

ORCID

Martin Klasen  <https://orcid.org/0000-0001-7823-3013>

REFERENCES

- Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews. Neuroscience*, 4(3), 165–178. <https://doi.org/10.1038/nrn1056>
- Alcalá-López, D., Smallwood, J., Jefferies, E., Van Overwalle, F., Vogele, K., Mars, R. B., ... Bzdok, D. (2018). Computing the social brain connectome across systems and states. *Cerebral Cortex*, 28(7), 2207–2232. <https://doi.org/10.1093/cercor/bhx121>
- Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*, 166(3–4), 559–571. <https://doi.org/10.1007/s00221-005-2396-5>
- Bowman, C. R., & Zeithamova, D. (2018). Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *The Journal of Neuroscience*, 38, 2605–2614. <https://doi.org/10.1523/jneurosci.2811-17.2018>
- Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, 8(4), 383–403. <https://doi.org/10.1016/j.plev.2011.10.002>
- Bzdok, D., Langner, R., Hoffstaedter, F., Turetsky, B. I., Zilles, K., & Eickhoff, S. B. (2012). The modular neuroarchitecture of social judgments on faces. *Cerebral Cortex*, 22(4), 951–961. <https://doi.org/10.1093/cercor/bhr166>
- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews. Neuroscience*, 6(8), 641–651. <https://doi.org/10.1038/nrn1724>
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, 11(12), 535–543. <https://doi.org/10.1016/j.tics.2007.10.001>
- Choi, I., Lee, J. Y., & Lee, S. H. (2018). Bottom-up and top-down modulation of multisensory integration. *Current Opinion in Neurobiology*, 52, 115–122. <https://doi.org/10.1016/j.conb.2018.05.002>
- Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, 16(2), 409–412. <https://doi.org/10.3758/bf03203962>
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428.
- Cona, G., & Semenza, C. (2017). Supplementary motor area as key structure for domain-general sequence processing: A unified account. *Neuroscience and Biobehavioral Reviews*, 72, 28–42. <https://doi.org/10.1016/j.neubiorev.2016.10.033>
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, 3(3), 201–215. <https://doi.org/10.1038/nrn755>
- Cox, D. E., & Harrison, D. W. (2008). Models of anger: Contributions from psychophysiology, neuropsychology and the cognitive behavioral perspective. *Brain Structure and Function*, 212(5), 371–385. <https://doi.org/10.1007/s00429-007-0168-7>
- Critchley, H. D., & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron*, 77(4), 624–638. <https://doi.org/10.1016/j.neuron.2013.02.008>
- Dekowska, M., Kuniecki, M., & Jaśkowski, P. (2008). Facing facts: Neuronal mechanisms of face perception. *Acta Neurobiologiae Experimentalis*, 68(2), 229–252.
- Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., ... Wildgruber, D. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport*, 17(3), 249–253. <https://doi.org/10.1097/01.wnr.0000199466.32036.5d>
- Ethofer, T., Bretschner, J., Wiethoff, S., Bisch, J., Schlipf, S., Wildgruber, D., & Kreifelts, B. (2013). Functional responses and structural connections of cortical areas for processing faces and voices in

- the superior temporal sulcus. *NeuroImage*, 76, 45–56. <https://doi.org/10.1016/j.neuroimage.2013.02.064>
- Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, 21(7), 1255–1268. <https://doi.org/10.1162/jocn.2009.21099>
- Frühholz, S., Hofstetter, C., Cristinzio, C., Saj, A., Seeck, M., Vuilleumier, P., & Grandjean, D. (2015). Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proceedings of the National Academy of Sciences of the United States of America*, 112(5), 1583–1588. <https://doi.org/10.1073/pnas.1411315112>
- Gawronski, B. (2002). What does the implicit association test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology*, 49(3), 171–180. <https://doi.org/10.1026/1618-3169.49.3.171>
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience*, 14(5), 350–363. <https://doi.org/10.1038/nrn3476>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- Harry, B., Williams, M. A., Davis, C., & Kim, J. (2013). Emotional expressions evoke a differential response in the fusiform face area. *Frontiers in Human Neuroscience*, 7, 1–6. <https://doi.org/10.3389/fnhum.2013.00692>
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, 51(1), 59–67.
- Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—It's my area: Or is it? *Journal of Cognitive Neuroscience*, 20(12), 2125–2136. <https://doi.org/10.1162/jocn.2008.20148>
- Hensel, L., Bzdok, D., Müller, V. I., Zilles, K., & Eickhoff, S. B. (2015). Neural correlates of explicit social judgments on vocal stimuli. *Cerebral Cortex*, 25(5), 1152–1162. <https://doi.org/10.1093/cercor/bht307>
- Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3(1), 80–84. <https://doi.org/10.1038/71152>
- Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, 3(3), 284–291. <https://doi.org/10.1038/72999>
- Hughes, S. M., Dispenza, F., & Gallup, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, 25(5), 295–304. <https://doi.org/10.1016/j.evolhumbehav.2004.06.001>
- Joassin, F., Maurage, P., & Campanella, S. (2011). The neural network sustaining the crossmodal processing of human gender from faces and voices: An fMRI study. *NeuroImage*, 54(2), 1654–1661. <https://doi.org/10.1016/j.neuroimage.2010.08.073>
- Klasen, M., Chen, Y. H., & Mathiak, K. (2012). Multisensory emotions: Perception, combination and underlying neural processes. *Reviews in the Neurosciences*, 23(4), 381–392. <https://doi.org/10.1515/revneuro-2012-0040>
- Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T., & Mathiak, K. (2011). Supramodal representation of emotions. *The Journal of Neuroscience*, 31(38), 13635–13643. <https://doi.org/10.1523/JNEUROSCI.2833-11.2011>
- Klasen, M., von Marschall, C., Isman, G., Zvyagintsev, M., Gur, R. C., & Mathiak, K. (2018). Prosody production networks are modulated by sensory cues and social context. *Social Cognitive and Affective Neuroscience*, 13(4), 418–429. <https://doi.org/10.1093/scan/nsy015>
- Knutson, K. M., Mah, L., Manly, C. F., & Grafman, J. (2007). Neural correlates of automatic beliefs about gender and race. *Human Brain Mapping*, 28(10), 915–930. <https://doi.org/10.1002/hbm.20320>
- Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, 86(3), 366–376.
- Latinus, M., VanRullen, R., & Taylor, M. J. (2010). Top-down and bottom-up modulation in processing bimodal face/voice stimuli. *BMC Neuroscience*, 11(1), 36. <https://doi.org/10.1186/1471-2202-11-36>
- Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience*, 6(10), 1108–1112. <https://doi.org/10.1038/nn1121>
- LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology*, 23(4), 727–738. <https://doi.org/10.1023/a:1025048802629>
- Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, 3(2), 215–221. <https://doi.org/10.3758/bf03212421>
- Mayer, A. R., Teshiba, T. M., Franco, A. R., Ling, J., Shane, M. S., Stephen, J. M., & Jung, R. E. (2012). Modeling conflict and error in the medial frontal cortex. *Human Brain Mapping*, 33(12), 2843–2855. <https://doi.org/10.1002/hbm.21405>
- Mazur, J. E. (2017). *Learning and behavior* (8th ed.). New York, NY: Routledge.
- McKay, R., Arciuli, J., Atkinson, A., Bennett, E., & Pheils, E. (2010). Lateralisation of self-esteem: An investigation using a dichotically presented auditory adaptation of the Implicit Association Test. *Cortex*, 46(3), 367–373.
- Menon, V. (2015). Saliency network. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference* (Vol. 2, pp. 597–611). Academic Press: Elsevier.
- Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure & Function*, 214(5–6), 655–667. <https://doi.org/10.1007/s00429-010-0262-0>
- Milne, E., & Grafman, J. (2001). Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 21, RC150. <https://doi.org/10.1523/JNEUROSCI.21-12-j0001.2001>
- Morris, J. S., Ohman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, 393(6684), 467–470. <https://doi.org/10.1038/30976>
- Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 108(24), 9998–10003. <https://doi.org/10.1073/pnas.1102433108>
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, 25(3), 653–660. <https://doi.org/10.1016/j.neuroimage.2004.12.005>
- Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, 220(3–4), 319–333. <https://doi.org/10.1007/s00221-012-3140-6>
- Ramsey, R., Cross, E. S., & Hamilton, A. F. (2013). Supramodal and modality-sensitive representations of perceived action categories in the human brain. *Experimental Brain Research*, 230(3), 345–357. <https://doi.org/10.1007/s00221-013-3659-1>
- Roitblat, H. L. (1987). *Introduction to comparative cognition*. New York, NY: W H Freeman/Times Books/Henry Holt & Co.
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, 16(2), 235–239. <https://doi.org/10.1016/j.conb.2006.03.001>
- Schirmer, A. (2018). Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing. *Social Cognitive and Affective Neuroscience*, 13(1), 1–13. <https://doi.org/10.1093/scan/nsx142>
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... Greicius, M. D. (2007). Dissociable intrinsic connectivity

- networks for salience processing and executive control. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(9), 2349–2356. <https://doi.org/10.1523/JNEUROSCI.5587-06.2007>
- Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D., & Warren, D. E. (2015). Investigating the neural correlates of schemas: Ventromedial prefrontal cortex is necessary for normal schematic influence on memory. *The Journal of Neuroscience*, 35(47), 15746–15751. <https://doi.org/10.1523/JNEUROSCI.2767-15.2015>
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Thonnessen, H., Boers, F., Dammers, J., Chen, Y. H., Norra, C., & Mathiak, K. (2010). Early sensory encoding of affective prosody: Neuromagnetic tomography of emotional category changes. *NeuroImage*, 50(1), 250–259. <https://doi.org/10.1016/j.neuroimage.2009.11.082>
- Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences*, 1124(1), 208–224. <https://doi.org/10.1196/annals.1440.012>
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., ... Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168(3), 242–249. <https://doi.org/10.1016/j.psychres.2008.05.006>
- Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews. Neuroscience*, 16(1), 55–61. <https://doi.org/10.1038/nrn3857>
- van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2007). Top-down task effects override automatic multisensory responses to letter-sound pairs in auditory association cortex. *NeuroImage*, 36(4), 1345–1360.
- Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, 9(12), 585–594. <https://doi.org/10.1016/j.tics.2005.10.011>
- Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, 30(3), 829–841. [https://doi.org/10.1016/S0896-6273\(01\)00328-2](https://doi.org/10.1016/S0896-6273(01)00328-2)
- Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, 7(11), 1271–1278. <https://doi.org/10.1038/nn1341>
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, 156, 249–268. [https://doi.org/10.1016/s0079-6123\(06\)56013-3](https://doi.org/10.1016/s0079-6123(06)56013-3)
- Wilson, M. C., & Scior, K. (2013). Attitudes towards individuals with disabilities as measured by the implicit association test: A literature review. *Research in Developmental Disabilities*, 35(2), 294–321.
- Woo, C. W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage*, 91, 412–419. <https://doi.org/10.1016/j.neuroimage.2013.12.058>
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association*, 310(20), 2191–2194. <https://doi.org/10.1001/jama.2013.281053>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Sonderfeld M, Mathiak K, Häring GS, et al. Supramodal neural networks support top-down processing of social signals. *Hum Brain Mapp*. 2021;42: 676–689. <https://doi.org/10.1002/hbm.25252>