Valence-specific conflict moderation in the dorso-medial PFC and the caudate head in emotional speech

Sonja A. Kotz,[1,2] Reinhard Dengler,[3] and Matthias Wittfoth[4,5]

[1]Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany, [2]School of Psychological Sciences, Zochonis Building, The University of Manchester, Brunswick Street, Manchester M13 9PL, UK, [3]Department of Neurology and Clinical Neurophysiology, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany, [4]Department of Psychiatry, Social psychiatry and Psychotherapy, Hannover Medical School Carl-Neuberg-Str. 1, 30625 Hannover, Germany, and [5]NICA – NeuroImaging and Clinical Applications, Hannover Medical School Carl-Neuberg-Str. 1, 30625 Hannover, Germany

Emotional speech comprises of complex multimodal verbal and non-verbal information that allows deducting others' emotional states or thoughts in social interactions. While the neural correlates of verbal and non-verbal aspects and their interaction in emotional speech have been identified, there is very little evidence on how we perceive and resolve incongruity in emotional speech, and whether such incongruity extends to current concepts of task-specific prediction errors as a consequence of unexpected action outcomes ('negative surprise'). Here, we explored this possibility while participants listened to congruent and incongruent angry, happy or neutral utterances and categorized the expressed emotions by their verbal (semantic) content. Results reveal valence-specific incongruity effects: negative verbal content expressed in a happy tone of voice increased activation in the dorso-medial prefrontal cortex (dmPFC) extending its role from conflict moderation to appraisal of valence-specific conflict in emotional speech. Conversely, the caudate head bilaterally responded selectively to positive verbal content expressed in an angry tone of voice broadening previous accounts of the caudate head in linguistic control to moderating valence-specific control in emotional speech. Together, these results suggest that control structures of the human brain (dmPFC and subcompartments of the basal ganglia) impact emotional speech differently when conflict arises.

INTRODUCTION

In social communication we often rely on verbal (e.g. semantic content) and non-verbal emotion expressions (e.g. tone of voice, face, gesture and body) to infer how someone feels and thinks. Most of these social exchanges employ congruent verbal and non-verbal emotion expressions that are known to facilitate information processing (e.g. de Gelder et al., 2004; Paulmann and Pell, 2009; Jessen and Kotz, 2011, 2013). However, communicating emotions is a complex process and, at times, marked by incongruity that leads to problems in decoding what a speaker intends to say (e.g. Schirmer et al., 2004; Mitchell, 2006a,b; Wittfoth et al., 2010). For example, in a professional conversation you generally do not expect someone to raise their voice in anger. However, a speaker, who is unhappy about the course of a conversation may smile at you, while her/his tone of voice may signal the emotion that the speaker really wants to express at that moment. This leads to mixed communicative signals and thus conflict in how to interpret the speaker's communicative intent. Consequently, it is not only important to understand how we: (i) perceive and integrate different forms of emotion expressions, but also (ii) perceive and resolve mixed communicative signals. We therefore need to identify the neural correlates that underlie the perception of verbal and non-verbal emotion expression when they do and do not match our expectations.

Previous functional magnetic resonance imaging (fMRI) research has identified predominantly right-lateralized fronto-temporal neural correlates for verbal and non-verbal congruent vocal emotion expressions (e.g. Bruck et al., 2011; Kotz and Paulmann, 2011; Paulmann et al., 2010). Furthermore, behavioral and electroencephalogram (EEG) patient studies have implicated the basal ganglia in the processing of non-verbal vocal emotion expressions such as tone of voice (e.g. Cancelliere and Kertesz, 1990; Pell, 1996; Pell and Leonard, 2003; Sidtis and Van Lancker Sidtis, 2003; Dara et al., 2008; Kotz et al., 2009; Paulmann and Pell, 2009; Paulmann et al., 2011), verbal emotion expressions (Castner et al., 2007; Hillier et al., 2007; Kotz et al., 2009; Paulmann and Pell, 2009) and the integration of emotion expressions (Paulmann and Pell, 2009; Paulmann et al., 2011). To date, only a few fMRI studies have reported enhanced activation of the basal ganglia in response to verbal vocal (Kotz and Paulmann, 2011) and non-verbal vocal emotion expressions (Morris et al., 1999; Wildgruber et al., 2002; Kotz et al., 2003, 2006; Grandjean et al., 2006; Bach et al., 2008; Fruhholz et al., 2012).

In an attempt to specify how verbal and non-verbal vocal emotion expressions interact, recent imaging studies have started to utilize incongruity paradigms (Schirmer et al., 2004; Mitchell, 2006a,b; Wittfoth et al., 2010). Results confirm the involvement of the above described fronto-striato-temporal network, but extend it to areas responding to incongruity or conflict: (i) the dorsal anterior cingulate cortex (dACC), more generally ascribed to conflict resolution (e.g. Kanske and Kotz, 2011a,b) and (ii) valence-specific responses in the caudate nucleus/thalamus (Wittfoth et al., 2010). In light of recent seminal work on cognitive and affective conflict that has implicated the role of the dACC/medial prefrontal cortex (mPFC) in monitoring predictions of action outcomes in specific task contexts (Alexander and Brown, 2011; Egner, 2011), the question arises whether a mixed communicative signal (e.g. an incongruent verbal and non-verbal emotion expression) reflects an unexpected action outcome (e.g. a speaker's communicative intent) that leads to a task-specific prediction error and increased

activation in dACC, mPFC or both (e.g. Egner, 2011). Further, it is of specific interest whether task-specific prediction errors, termed 'negative surprise' (Egner, 2011) are valence-specific or not (Alexander and Brown, 2011). Taken together, such evidence would indeed confirm (i) a domain-general role of the dACC/mPFC in responding to prediction errors independent of the stimulus type used to induce conflict in task-specific action outcome and (ii) would specify whether 'negative surprise' is a valence-specific phenomenon or not.

Therefore, we set out to test and extend the following research questions: (i) do dACC, mPFC or both respond to a task-specific prediction error (e.g. incongruent verbal and non-verbal emotion expression)? and (ii) is the response in these areas valence-specific or not?

We instructed participants to categorize the emotional content of an utterance (task specificity: emotional semantic content). We expected to find activation in dACC, mPFC or both in response to incongruent verbal and non-verbal emotion expressions. Should the unexpected outcome of mismatching verbal and non-verbal expressions be valence-specific, we predicted activation in dACC, mPFC or both to increase when negative verbal content is spoken in a happy tone of voice (task-specific 'negative surprise'). On the other hand, a mismatch between positive verbal content and an angry tone of voice may engage subcomponents of the basal ganglia (e.g. caudate; Wittfoth *et al.*, 2010) in support of a valence-specific response to an unexpected action outcome in emotional speech.

MATERIALS AND METHODS

Participants

In total, 20 right-handed, normal hearing native speakers of German were recruited from the University of Magdeburg student population with an age range of 24–37 years (mean age 26.9 years; s.d. = 3.4, 10 females). None of them reported any current or past medical or neurological condition. They received a brief MR safety screening and gave informed consent prior to the experiment. The protocol was approved by the local ethics committee of the University of Magdeburg and of the Hannover Medical School in accordance with the Declaration of Helsinki.

Paradigm

The stimulus material consisted of 96 German sentences that were spoken by a semi-professional actress. Stimuli were recorded with a Digital audio tape (DAT) recorder, digitized at a 16 bit/44.1 kHz sampling rate, and normalized to 75 dB. The material consisted of an equal number (32) of three verbal ($_v$) emotion expressions: positive (P) (e.g. 'Sie hat ihre Prüfung bestanden.'; 'She has passed her exam.'), negative (N) (e.g. 'Sie hat ihm das Herz gebrochen.'; 'She has broken his heart.'), or neutral (X) ('Sie hat das Buch gelesen.'; 'She has read the book.'). Each of the three verbal emotion expressions was spoken either with a positive (happy), negative (angry) or neutral tone of voice ($_{nv}$) resulting in a total number of 288 experimental stimuli. Of the resulting nine experimental conditions, three were congruent [congruent positive (P_vP_{nv})]; congruent negative (N_vN_{nv}); congruent neutral (X_vX_{nv})], whereas two were incongruent [incongruent positive (P_vN_{nv}); incongruent negative (N_vP_{nv})]. The remaining conditions were paired with neutral verbal (X_vP_{nv}, X_vN_{nv}) or non-verbal vocal expressions (P_vX_{nv}, N_vX_{nv}) and served as filler conditions to counterbalance the design. The average duration of an expression was about 1.6 s. Of further note is that the experimental stimuli were constructed in a way that verbal emotion expression could only be recognized at the end of an utterance (the noun), whereas the impact of non-verbal vocal information started with the onset of an utterance (Paulmann and Kotz, 2008). Thus, we aimed at studying the interaction of non-

verbal vocal and verbal emotion expressions in a semi-controlled fashion.

Participants categorized the verbal emotional content of each expression by pressing one of three buttons on a custom-built response box with their right hand as fast and accurately as possible. Half of the participants responded with the right-hand index finger to positive verbal content and with their right-hand ring finger to negative verbal content; the response assignment was reversed for the other half of the participants. All participants pressed with the right-hand middle finger in response to neutral verbal content. Auditory stimuli were delivered using a high-frequency shielded transducer system. This transmission system includes a piezoelectric loudspeaker enabling the transmission of strong sound pressure levels (∼105 dB) with excellent attenuation characteristics. The loudspeakers were embedded in tightly occlusive MR-compatible headphones allowing unimpeded conduction of the stimuli. Participants wore noise protection ear plugs to provide additional noise attenuation. Auditory stimulation was adjusted during a short training session to a comfortable listening level, which was the same for all participants. After the scanning, all participants reported that the acoustic quality and comprehension of the presented stimuli was optimal.

MRI data acquisition

Images were acquired employing a standard head coil with a Siemens Trio 3-T scanner (Erlangen, Germany) at the ZENIT (Zentrum für neurowissenschaftliche Innovation und Technologie, University Clinic of Magdeburg, Germany). Functional images were recorded axially along the AC-PC plane with a T2*-weighted gradient-echo echo-planar imaging (EPI) sequence (repetition time, TR = 3500 ms; echo time, TE = 30 ms; flip angle = 90°, FoV 192×192 mm^2, matrix 64×64, $3 \times 3 \times 5$ mm in-plane resolution) with 18 slices of 4 mm thickness interspersed with 1 mm in-between gaps (seven slices extended ventrally from the AC-PC plane). Slices were acquired in an interleaved manner. One session was recorded with two short breaks of stimulation (2×20 s, no scanner interruption), resulting in a total of 658 volumes. All participants were instructed to lie still during the stimulation and pauses and to shut their eyes throughout the scanning session. The first three volumes were discarded to allow for magnetic saturation effects. For the structural scans, we used a T1-weighted sequence in the same orientation as the functional sequence to provide detailed anatomical images aligned to the functional scans (192 1 mm slices, TR 2500 ms, TE 4.77 ms, FoV of 256 mm). High-resolution structural images were also acquired for the purpose of cross-subject registration. Scanning consisted of an event-related design with jittered inter-trial-intervals randomly chosen from a distribution ranging from 0 to 2 s. About 48 null events were included in the design. All stimuli were presented in pseudo-randomized order with the constraint that no expression type was immediately repeated in the same condition.

Data analysis

'Behavioral data' acquired during scanning were assessed with respect to response times and error rates. A two-factorial repeated-measures analysis of variance (ANOVA) with the factors verbal emotion expression (positive, negative and neutral) and non-verbal vocal emotion expression (happy, angry and neutral tone of voice) was conducted for reaction times and error rates. The statistical threshold was set to $P < 0.05$ and Greenhouse–Geisser corrections were applied when necessary.

'Functional MRI data' were pre-processed with Statistical Parametric Mapping (SPM5) (Welcome Department of Cognitive Neurology, London). As the slices of each volume were not acquired simultaneously, a timing correction procedure was applied. All

volumes were realigned to the 10th volume, 'unwarped' to remove variance caused by movement-by-field in homogeneity interactions, normalized to a standard EPI template and smoothed with a Gaussian kernel of 8 mm full-width at half-maximum to account for anatomical differences between participants and to allow for statistical inference using the Gaussian Random Field theory. At the first level, data were high-pass filtered (128 s) and an autoregressive function (AR-1) was employed to estimate temporal autocorrelation in the data and to correct the degrees of freedom, accordingly.

The first-level model consisted of 10 regressors, which comprised all experimental conditions and an additional regressor, in which incorrect or delayed response trials were entered. Trial onsets in SPM were time points at the end of each expression motivated by the expression characteristics (early and on-going vocal emotion expression and late verbal emotion expression). As mentioned above, expressions were composed in a way that participants were not able to identify verbal emotional content before hearing the last word of the utterance. This ensured that participants fully processed the spoken utterance before categorizing the verbal content of an utterance.

Nine regressors (all experimental conditions) were analysed in a repeated-measure ANOVA. In order to examine incongruity effects, we restricted our analyses to those conditions, which did not contain neutral information (see explanation above). As the task-specific feature was always the verbal content of an expression, the task-irrelevant information (non-verbal vocal emotion expressions) was either congruent (P_vP_{nv}, N_vN_{nv}) or incongruent (P_vN_{nv}, N_vP_{nv}).

In order to address our specific research questions, we report (i) an overall incongruity effect, (ii) two task-specific incongruity effects by comparing incongruent and congruent expressions when the verbal emotion expressions (task-specific) changed valence ($P_vN_{nv} > N_vN_{nv}$ and $N_vP_{nv} > P_vP_{nv}$), but non-verbal vocal emotion expressions (task-irrelevant) stayed constant and (iii) valence-specific effects (congruent).

We applied a statistical threshold of $P < 0.001$ with an additional cluster extend threshold corrected for multiple comparisons in order to protect against false positive results (Forman *et al.*, 1995). This threshold ($k = 13$ voxels) was calculated by SPM using an empirically determined extent threshold. This method is implemented in SPM and refers to the estimated smoothness of the images. After determining of the number of resels, the expected Euler characteristic is calculated. This is used to give the correct threshold (number of voxels) that is required to control for false positive results. Coordinates of cluster maxima are reported in MNI space and labeled using the aal database within the WFU-Pickatlas (Tzourio-Mazoyer *et al.*, 2002; Maldjian *et al.*, 2003, 2004).

RESULTS

Behavioral data

Main effects for both verbal and non-verbal information were observed when analysing reaction times (RTs) and error rates (ERs).

Participants responded slowest to positive verbal expressions and fastest to negative verbal expressions [main effect verbal content for RT: $F(2, 38) = 44.7$, $P < 0.001$]. However, participants displayed lower ERs for positive verbal expressions and increased ERs for neutral verbal expressions [main effect verbal for ER: $F(2, 38) = 7.04$, $P = 0.003$].

When sentences were neutrally intoned, participants responded slower compared to sentences spoken in an emotional tone of voice [main effect non-verbal for RT: $F(2, 38) = 22.63$, $P < 0.001$]. Further, participants made more errors responding to sentences spoken in a neutral tone of voice than to sentences in a happy tone of voice [main effect non-verbal for ER: $F(2, 38) = 4.61$, $P = 0.038$].
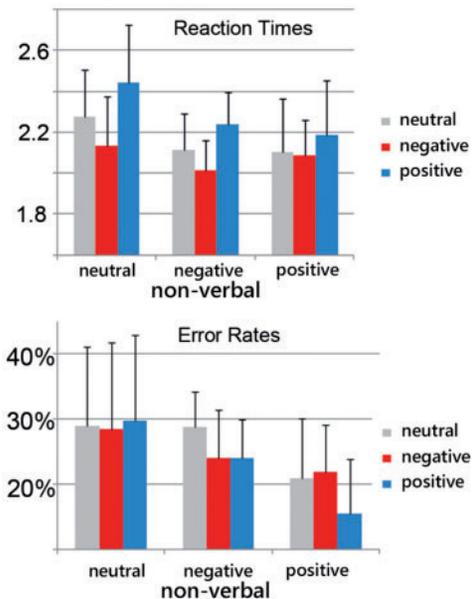

Fig. 1 Behavioral data: average reaction times and error rates. Reaction times in response to verbal emotion expressions were slowest responding to positive verbal information and fastest responding to negative verbal information. Furthermore, participants responded fastest when listening to congruent negative expressions (N_vN_{nv}).

For both reaction times $F(4, 76) = 6.3$; $P < 0.001$) and accuracy rates [$F(4, 76) = 4.4$; $P = 0.003$], we also found an interaction of verbal and non-verbal emotion expressions (Figure 1).

Resolving this interaction with Bonferroni-corrected post hoc *t*-tests ($P = 0.05/4 = 0.0125$) for RTs, we found that participants responded significantly faster to congruent negative sentences (N_vN_{nv}) compared with both incongruent conditions (vs. P_vN_{nv}: t(19) = −10,8, p < 0.001; vs. N_vP_{nv}: t(19) = −3,27, p = 0.004). Comparing positive congruent sentences with both incongruent conditions did only yielded one significant result (*vs* N_vP_{nv}: $P = 0.016$, *vs* P_vN_{nv}: $P = 0.171$).

With respect to ERs, we found that participants made less errors in the congruent positive condition (P_vP_{nv}) compared with both incongruent conditions [*vs* N_vP_{nv}: $t(19) = 5.12$, $P = 0.00006$, *vs* P_vN_{nv}: $t(19) = 4.22$, $P = 0.00047$]. Here, comparing congruent negative sentences (N_vN_{nv}) with the two incongruent conditions did not yield significant results (*vs* P_vN_{nv}: $P = 1.0$, *vs* N_vP_{nv}: $P = 0.29154$).

Functional MRI data

Overall incongruity

As listed in Table 1, attenuated activation in the right thalamus was found in the overall contrast between incongruent and congruent utterances.

Task (verbal)- and valence-specific incongruity effects

When compared with congruent negative expressions, positive verbal expressions spoken in an angry tone of voice ($P_vN_{nv} > N_vN_{nv}$) revealed increased activation in the caudate head bilaterally, the left hippocampus, the right precuneus and the right thalamus. Negative verbal expressions intoned happily compared with congruent positive expressions ($N_vP_{nv} > P_vP_{nv}$) increased activation in the bilateral posterior superior temporal gyrus (STG), the right mid-STG, the right middle frontal gyrus (MFG; BA46) and in the left inferior frontal gyrus (IFG; BA47; Table 1).

Table 1 Activation clusters of task-specific incongruity effects

Region	Side	BA	z-value	Cluster size	MNI coordinates		
					X	Y	Z
Incongruent ($P_vN_{nv} + N_vP_{nv}$) > Congruent ($P_vP_{nv} + N_vN_{nv}$)							
Thalamus	R		4.29	59	2	−18	22
Incongruent (P_vN_{nv}) > Congruent (N_vN_{nv})							
Caudate head	L		4.30	102	−10	20	2
	R		3.88	110	10	20	4
Hippocampus	L		4.11	73	−32	−44	0
Thalamus	R		3.99	51	2	−18	22
Precuneus	R		3.84	23	26	−48	20
Incongruent (N_vP_{nv}) > Congruent (P_vP_{nv})							
Mid-STG	R	22	3.77	113	70	−26	4
Posterior STG	R	22	3.65	19	50	−36	4
	L	22	3.54	20	−62	−32	8
MFG	R	46	3.58	13	46	30	20
IFG	L	47	3.48	16	−46	22	4

Order of presentation: overall incongruity, positive conflict, negative conflict.
Whole-brain analysis of all suprathreshold voxels; contrasts at $P < 0.001$, $k = 13$.
SFG, superior frontal gyrus.

Table 2 Activation clusters of task- and valence-specific effects

Region	Side	BA	z-value	Clustersize	MNI coordinates		
					X	Y	Z
Positive > negative verbal expressions							
Lingual gyrus	R	19	4.71	187	38	−58	0
	L	19	3.46	37	−28	−64	−2
Caudate head	R		4.24	98	10	20	2
	L		3.76	32	−10	20	4
Negative > positive verbal expressions							
dmPFC	L	10	4.70	301	−6	52	22
	R	10	4.21	45	20	50	6
Mid-MTG	R	21	4.49	283	66	−26	−14
ITG	L	20	4.03	34	−52	−2	−30
Posterior STG	L	22/42	3.82	34	−62	−32	8
	L	41	3.44	15	−40	−38	12
SFG	R	10	3.36	15	20	56	26

Whole-brain analysis of all suprathreshold voxels; contrasts at $P < 0.001$, $k = 13$.
SFG, superior frontal gyrus.

Task (verbal)- and valence-specific effects

Contrasting congruent positive (PvPnv) and negative (NvNnv) emotion expressions revealed increased activation in the caudate head bilaterally and the lingual gyri (BA19). The reversed contrast led to increased activation in the dorso-medial PFC (dmPFC) bilaterally (BA10), the left inferior temporal gyrus (ITG; BA20), the right middle temporal gyrus (MTG; BA 21) and left posterior STG (BA22/41/42) (Table 2 and Figure 2).

DISCUSSION

In this study we set out to investigate whether incongruity of verbal (positive, negative and neutral) and non-verbal vocal (happy, angry and neutral) emotion expressions leads to activation in dACC, mPFC or both comparable with reported activation in response to task-specific prediction errors of expected action outcomes (Alexander and Brown, 2011; Egner, 2011) and whether such activation is valence-specific or not (Wittfoth et al., 2010; Alexander and Brown, 2011).

In a fully crossed design, participants categorized the verbal emotional content of utterances (task specificity) they listened to.

Overall incongruity-related activation

To test how verbal and non-verbal vocal emotion expressions interact, we applied an incongruity paradigm. Unlike the Stroop task, in which a neutral stimulus consists of a color-word written in an ink color unrelated to it, emotional verbal expressions spoken in an unexpected tone of voice can be informative in social communicative interactions (i.e. stepping over social boundaries in a communicative context). It is important to note that the expressions used in this experiment ensured that participants were only able to detect a possible incongruity of verbal and non-verbal vocal emotional expressions at the end of an utterance. Our finding that the thalamus is associated with an incongruity response, is in line with the view that this structure is part of a network, which comes into play during the anticipation of aversive events (Chua et al., 1999; Ploghaus et al., 1999) as well as subsequent emotion regulation with a possible impact on conflict resolution (Herwig et al., 2007; Goldin et al., 2008). Furthermore, the thalamus responded specifically to an angry tone of voice that did not match the verbal emotional content, thereby emphasizing its role as a threat detector (Choi et al., 2012).

Task- and valence-specific effects

When comparing positive verbal emotion expressions intoned angrily, we observed a task-specific (verbal content) activation increase in the caudate heads and in the lingual gyri. This pattern of activation is comparable with reported previous results (Wittfoth et al., 2010), which revealed increased activation of the caudate heads when positive verbal expressions were spoken in an angry tone of voice (task specificity: tone of voice). So, independent of task requirements we confirm a valence-specific activation pattern. Further, activation of the caudate heads not only occurred in response to positive verbal expressions, but even more so when verbal expressions mismatched non-verbal emotional vocal expression. We suggest that the role of the caudate head(s) goes beyond the detection of positive social incentives, but also moderates the control of verbal and non-verbal vocal interaction (i.e. its congruity or incongruity) independent of task requirements (i.e. focus on verbal or non-verbal emotion expressions). On the one hand, this view is supported by observations that the caudate head is responsive to positive social signals in economic games (King-Casas et al., 2005; Delgado et al., 2008) or in the broader sense, whenever an outcome is perceived as desirable (Grahn et al., 2008). Our results may therefore extend monetary rewards to social communicative incentives. On the other hand, recent studies have emphasized the role of the caudate head in language control, i.e. during the processing of incongruent words (Ali et al., 2010) or language switching in bilinguals (Crinion et al., 2006). Furthermore, a recent meta-analytic connectivity study associated the head of the caudate nucleus with cognition and emotion-related structures (e.g. the IFG or the dACC), which seem to play a relevant role in the present investigation (Robinson et al., 2012). Crucially, independent of task demands [verbal here, non-verbal elsewhere (Wittfoth et al., 2010)] or the incongruity of an emotion expression, the caudate head responds in a valence-specific manner. Consequently, the present results may encourage investigations of possible structural or neurodegenerative changes of the caudate head (Paulmann et al., 2011; Robotham et al., 2011). Although there have been some attempts to link executive dysfunction to reduced activation in striatal regions in Parkinson's or Huntington's disease (Hasselbalch et al., 1992; Lawrence et al., 1998), the present finding may offer new insight into valence-specific incongruity effects in emotional speech and beyond in these patient groups.
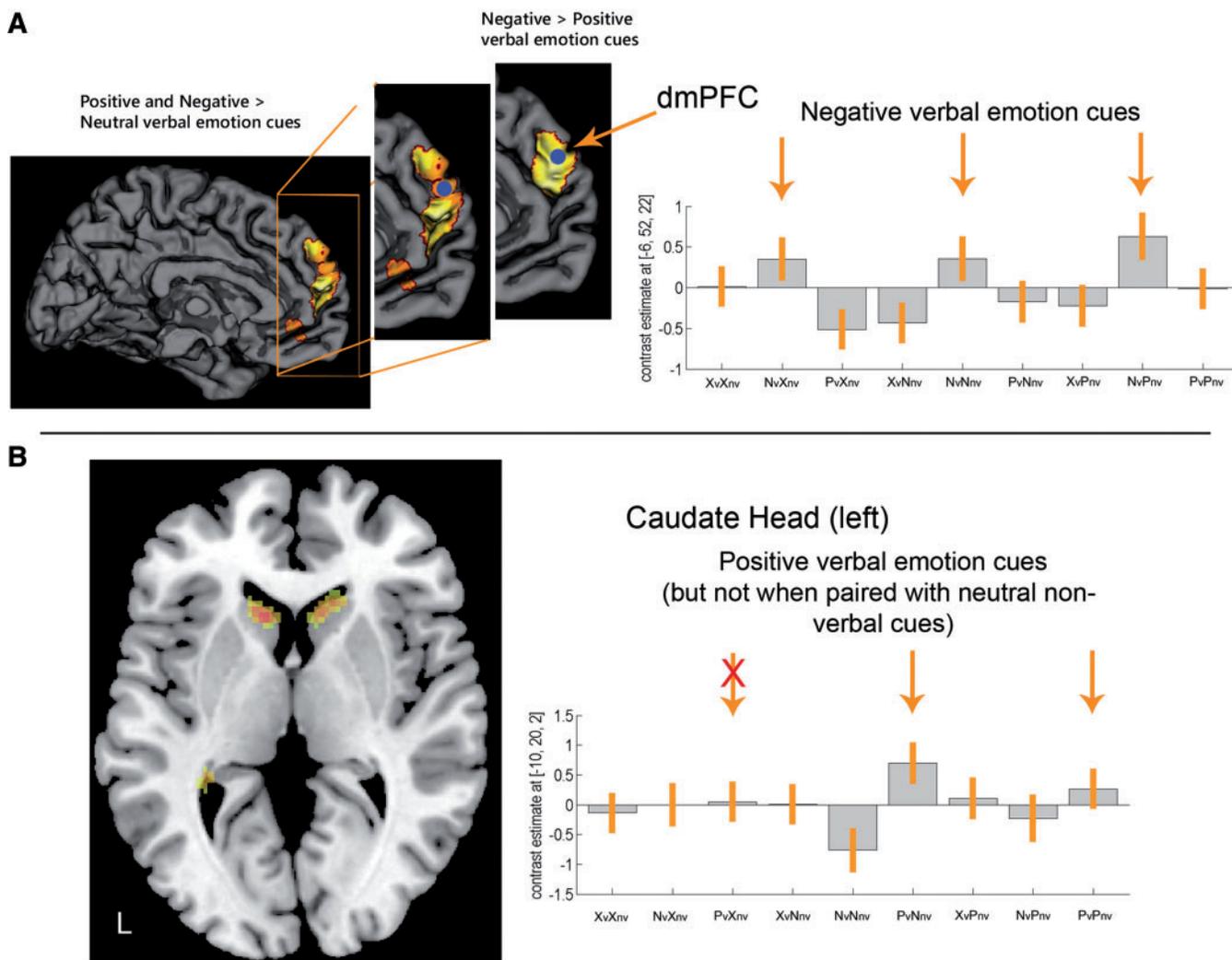
Fig. 2 Task- and valence-specific effects. (**A**) Displayed are significant clusters in the left dmPFC. Contrast estimates of the left dmPFC support the view that negative verbal expressions (cues) are processed by this area. Blue circles indicate the area from which the contrast estimates were extracted. (**B**) Significant clusters in bilateral caudate heads shown in an axial view (z = 2). Contrast estimates of the left caudate head displays elevated signals for positive verbal emotion expressions (cues). Contrast estimates for the right caudate head were comparable (data not shown here). Furthermore, activation of the left lingual gyrus can be seen.

Taking a closer look at the processing of negative compared with positive verbal emotion expressions, we find increased activation in the bilateral mid-to-posterior STG and in the dmPFC. In particular, the activation of the dmPFC is of interest as this region has been attributed to inferring other people's mental states (theory-of-mind) (Beaucousin *et al.*, 2007; Sebastian *et al.*, 2012). Furthermore, a recent imaging study using multi-voxel pattern analysis suggested not only a specific role of the dmPFC in mental state attribution, but found that this area together with the left superior temporal sulcus (STS) responds modality independently (Peelen *et al.*, 2010). Our findings corroborate this view of the dmPFC as a region associated with (i) appraisal as well as (ii) conflict regulation in humans (Etkin *et al.*, 2011) in a modality independent manner (Peelen *et al.*, 2010).

However, on the basis of these results, we propose that the role of the dmPFC extends to valence- and task-specific conflict regulation. Only negative verbal expression, but not positive verbal expression as well as the incongruity of negative verbal expression and happy tone of voice led to increased activation in the dmPFC. Additional activation was observed in the right MFG, an area within the lateral prefrontal cortex, which has frequently been associated with the processing of

semantic constraints in sentences comprehension [for an overview, see Price (2010)] or with the processing of incongruent non-verbal vocal expressions (Wittfoth *et al.,* 2010).

One crucial question remains to be discussed at this stage: why may dmPFC activation be specific for 'negative' conflict and the caudate for 'positive' conflict? More specifically, can these valence-specific activation patterns be considered as domain general effects, that is, not speech (verbal or non-verbal) specific? Upon elaborations on the valence-specific response of the caudate heads within the basal ganglia above, we will further consider the specific role of dmPFC in response to negative conflict.

Our finding of a distinction of a valence-specific response pattern in dmPFC is supported by a number of studies. These results indicate that dmPFC activation links to a number of different processes, none of them mutually exclusive. These processes include the detection of change in the environment that signal the subsequent need for behavioral adjustment, especially negative change including unfavorable outcomes and response conflict (Holroyd and Coles, 2002; Rushworth *et al.*, 2004; Yeung *et al.*, 2004; Botvinick, 2007; Alexander and Brown 2011), social threats (Eisenberger and Lieberman, 2004; Eisenberger, 2013) or emotional reappraisal (Etkin *et al.*, 2011;

Ochsner *et al.* 2012). In particular, recent investigations of the 'social brain' have suggested that dmPFC may respond sensitively to cognitive imbalance and may therefore reflect a degree of cognitive dissonance (Izuma *et al.*, 2010; Izuma and Adolphs 2013). In other words, dmPFC encodes the discrepancy of current negative information and a desired outcome.

Accumulating data also suggest that dmPFC activation corresponds to the ability to reason about other peoples' mental states, generally referred to as theory of mind (Ochsner *et al.*, 2004). Considering our experimental setup, in which participants listen to statements of one person about a third person, our findings fit well to the concept that dmPFC may come into play in triadic relations between two minds and an object (Saxe, 2006). Our results combine and can extend these perspectives of the role of the dmPFC. We therefore suggest that dmPFC encodes the social context of triadic relations of negative flavor in emotional speech, thus inducing an emotional imbalance and signaling a need for a subsequent adjustment.

CONCLUSIONS

This study set out to shed further light on the interaction of verbal and non-verbal vocal emotion expressions. Utilizing an incongruity paradigm it also allowed investigating whether a mismatch between verbal and non-verbal emotion expressions, as often to be found in social communicative interactions, relies on valence-specific activation patterns in the human brain that may or may not be domain specific. These results clearly emphasize that the appraisal and monitoring of emotional speech incongruities extend beyond classically described networks involved in the perception and the categorization of verbal and non-verbal vocal expressions. Both the dmPFC and the caudate head in their more general functions of conflict regulation and cognitive/affective control, respectively, respond in a valence-specific manner when encountering incongruity in emotional speech and confirm the importance of these control structures in human social communication.

Conflict of Interest

None declared.

REFERENCES

Alexander, W.H., Brown, J.W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, 14, 1338–44.

Ali, N., Green, D.W., Kherif, F., Devlin, J.T., Price, C.J. (2010). The role of the left head of caudate in suppressing irrelevant words. *Journal of Cognitive Neuroscience*, 22, 2369–86.

Bach, D.R., Grandjean, D., Sander, D., Herdener, M., Strik, W.K., Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *Neuroimage*, 42, 919–27.

Beaucousin, V., Lacheret, A., Turbelin, M.R., Morel, M., Mazoyer, B., Tzourio-Mazoyer, N. (2007). FMRI study of emotional speech comprehension. *Cerebral Cortex*, 17, 339–52.

Botvinick, M.M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive and Affective Behavioral Neuroscience*, 7(4), 356–66.

Bruck, C., Kreifelts, B., Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, 8, 383–403.

Cancelliere, A.E., Kertesz, A. (1990). Lesion localization in acquired deficits of emotional expression and comprehension. *Brain and Cognition*, 13, 133–47.

Castner, J.E., Chenery, H.J., Copland, D.A., Coyne, T.J., Sinclair, F., Silburn, P.A. (2007). Semantic and affective priming as a function of stimulation of the subthalamic nucleus in Parkinson's disease. *Brain*, 130, 1395–407.

Choi, J.M., Padmala, S., Pessoa, L. (2012). Impact of state anxiety on the interaction between threat monitoring and cognition. *Neuroimage*, 59, 1912–23.

Chua, P., Krams, M., Toni, I., Passingham, R., Dolan, R. (1999). A functional anatomy of anticipatory anxiety. *Neuroimage*, 9, 563–71.

Crinion, J., Turner, R., Grogan, A., et al. (2006). Language control in the bilingual brain. *Science*, 312, 1537–40.

Dara, C., Monetta, L., Pell, M.D. (2008). Vocal emotion processing in Parkinson's disease: reduced sensitivity to negative emotions. *Brain Research*, 1188, 100–11.

de Gelder, B., Snyder, J., Greve, D., Gerard, G., Hadjikhani, N. (2004). Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Science USA*, 101, 16701–6.

Delgado, M.R., Schotter, A., Ozbay, E.Y., Phelps, E.A. (2008). Understanding overbidding: using the neural circuitry of reward to design economic auctions. *Science*, 321, 1849–52.

Eisenberger, N.I. (2013). Social ties and health: a social neuroscience perspective. *Current Opinion in Neurobiology*, 23(3), 407–13.

Eisenberger, N.I., Lieberman, M.D. (2004). Why rejection hurts: a common neural alarm system for physical and social pain. *Trends in Cognitive Science*, 8(7), 294–300.

Egner, T. (2011). Surprise! A unifying model of dorsal anterior cingulate function? *Nature neuroscience*, 14, 1219–20.

Etkin, A., Egner, T., Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Science*, 15(2), 85–93.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33, 636–47.

Fruhholz, S., Ceravolo, L., Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, 22(5), 1107–17.

Goldin, P.R., McRae, K., Ramel, W., Gross, J.J. (2008). The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biological Psychiatry*, 63, 577–86.

Grahn, J.A., Parkinson, J.A., Owen, A.M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, 86, 141–55.

Grandjean, D., Banziger, T., Scherer, K.R. (2006). Intonation as an interface between language and affect. *Progress in Brain Research*, 156, 235–47.

Hasselbalch, S.G., Oberg, G., Sorensen, S.A., et al. (1992). Reduced regional cerebral blood flow in Huntington's disease studied by SPECT. *Journal of Neurology, Neurosurgery & Psychiatry*, 55, 1018–23.

Herwig, U., Baumgartner, T., Kaffenberger, T., et al. (2007). Modulation of anticipatory emotion and perception processing by cognitive control. *Neuroimage*, 37, 652–62.

Hillier, A., Beversdorf, D.Q., Raymer, A.M., Williamson, D.J., Heilman, K.M. (2007). Abnormal emotional word ratings in Parkinson's disease. *Neurocase*, 13, 81–5.

Holroyd, C.B., Coles, M.G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709.

Izuma, K., Adolphs, R. (2013). Social manipulation of preference in the human brain. *Neuron*, 78(3), 563–73.

Izuma, K., Matsumoto, M., Murayama, K., Samejima, K., Sadato, N., Matsumoto, K. (2010). Neural correlates of cognitive dissonance and choice-induced preference change. *Proceedings of the National Academy of Science USA*, 107(51), 22014–9.

Jessen, S., Kotz, S.A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage*, 58, 665–74.

Jessen, S., Kotz, S.A. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Frontiers in Human Neuroscience*, 7, 369.

Kanske, P., Kotz, S.A. (2011a). Conflict processing is modulated by positive emotion: ERP data from a flanker task. *Behavioural Brain Research*, 219(2), 382–6.

Kanske, P., Kotz, S.A. (2011b). Positive emotion speeds up conflict processing: ERP responses in an auditory Simon task. *Biological Psychology*, 87(1), 122–7.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, 308, 78–83.

Kotz, S.A., Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, 5(3), 108–25.

Kotz, S.A., Schwartze, M., Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex*, 45, 982–90.

Kotz, S.A., Meyer, M., Paulmann, S. (2006). Lateralization of emotional prosody in the brain: An overview and synopsis on the impact of study design. *Progress in Brain Research*, 156, 285–94.

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, 366–76.

Lawrence, A.D., Weeks, R.A., Brooks, D.J., et al. (1998). The relationship between striatal dopamine receptor binding and cognitive performance in Huntington's disease. *Brain*, 121(Pt 7), 1343–55.

Maldjian, J.A., Laurienti, P.J., Burdette, J.H. (2004). Precentral gyrus discrepancy in electronic versions of the Talairach atlas. *Neuroimage*, 21, 450–5.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*, 19, 1233–9.

Mitchell, R.L. (2006a). How does the brain mediate interpretation of incongruent auditory emotions? The neural response to prosody in the presence of conflicting lexico-semantic cues. *European Journal of Neuroscience*, 24, 3611–8.

Mitchell, R.L. (2006b). Does incongruence of lexicosemantic and prosodic information cause discernible cognitive conflict? *Cognitive, Affective & Behavioral Neuroscience*, 6, 298–305.

Morris, J.S., Scott, S.K., Dolan, R.J. (1999). Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia*, 37, 1155–63.

Ochsner, K.N., Knierim, K., Ludlow, D.H., et al. (2004). Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience*, 16(10), 1746–72.

Ochsner, K.N., Silvers, J.A., Buhle, J.H. (2012). Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, 1251, E1–24.

Paulmann, S., Kotz, S.A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language*, 105, 59–69.

Paulmann, S., Pell, M.D. (2009). Facial expression decoding as a function of emotional meaning status: ERP evidence. *Neuroreport*, 20, 1603–8.

Paulmann, S., Seifert, S., Kotz, S.A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience*, 5, 59–75.

Paulmann, S., Ott, D.V., Kotz, S.A. (2011). Emotional speech perception unfolding in time: the role of the Basal Ganglia. *PLoS One*, 6, e17694.

Peelen, M.V., Atkinson, A.P., Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, 30, 10127–34.

Pell, M.D. (1996). On the receptive prosodic loss in Parkinson's disease. *Cortex*, 32, 693–704.

Pell, M.D., Leonard, C.L. (2003). Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. *Cognitive, Affective & Behavioral Neuroscience*, 3, 275–88.

Ploghaus, A., Tracey, I., Gati, J.S., et al. (1999). Dissociating pain from its anticipation in the human brain. *Science*, 284, 1979–81.

Price, C.J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191, 62–88.

Robinson, J.L., Laird, A.R., Glahn, D.C., et al. (2012). The functional connectivity of the human caudate: An application of meta-analytic connectivity modeling with behavioral filtering. *Neuroimage*, 60, 117–29.

Robotham, L., Sauter, D.A., Bachoud-Levi, A.C., Trinkler, I. (2011). The impairment of emotion recognition in Huntington's disease extends to positive emotions. *Cortex*, 47, 880–4.

Rushworth, M.F., Walton, M.E., Kennerley, S.W., Bannerman, D.M. (2004). Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Science*, 8(9), 410–7.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, 16(2), 235–9.

Schirmer, A., Zysset, S., Kotz, S.A., Yves von Cramon, D. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *Neuroimage*, 21, 1114–23.

Sebastian, C.L., Fontaine, N.M., Bird, G., et al. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive & Affective Neuroscience*, 7(1), 53–63.

Sidtis, J.J., Van Lancker Sidtis, D. (2003). A neurobehavioral approach to dysprosody. *Seminars in Speech and Language*, 24, 93–105.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15, 273–89.

Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *Neuroimage*, 15, 856–69.

Wittfoth, M., Schroder, C., Schardt, D.M., Dengler, R., Heinze, H.J., Kotz, S.A. (2010). On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. *Cerebral Cortex*, 20, 383–92.

Yeung, N., Botvinick, M.M., Cohen, J.D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, 111(4), 931–59.