Social Cognitive and Affective Neuroscience, 2018, 604–615

doi: 10.1093/scan/nsy034
Advance Access Publication Date: 16 May 2018
Original article

Neural bases of social communicative intentions in speech

Nele Hellbernd and Daniela Sammler

Otto Hahn Group Neural Bases of Intonation in Speech and Music, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraβe 1a, D-04103 Leipzig, Germany

Correspondence should be addressed to Daniela Sammler, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany. E-mail: sammler@cbs.mpg.de

Abstract

Our ability to understand others' communicative intentions in speech is key to successful social interaction. Indeed, misunderstanding an 'excuse me' as apology, while meant as criticism, may have important consequences. Recent behavioural studies have provided evidence that prosody, that is, vocal tone, is an important indicator for speakers' intentions. Using a novel audio-morphing paradigm, the present functional magnetic resonance imaging study examined the neurocognitive mechanisms that allow listeners to 'read' speakers' intents from vocal prosodic patterns. Participants categorized prosodic expressions that gradually varied in their acoustics between criticism, doubt, and suggestion. Categorizing *typical* exemplars of the three intentions induced activations along the ventral auditory stream, complemented by amygdala and mentalizing system. These findings likely depict the stepwise conversion of external perceptual information into abstract prosodic categories and internal social semantic concepts, including the speaker's mental state. *Ambiguous* tokens, in turn, involved cingulo-opercular areas known to assist decision-making in case of conflicting cues. Auditory and decision-making processes were flexibly coupled with the amygdala, depending on prosodic typicality, indicating enhanced categorization efficiency of overtly relevant, meaningful prosodic signals. Altogether, the results point to a model in which auditory prosodic categorization and socio-inferential conceptualization cooperate to translate perceived vocal tone into a coherent representation of the speaker's intent.

Key words: voice; prosody; intention; theory of mind; auditory categorical perception; connectivity

Introduction

Successful social interaction requires a minimal understanding of others' intentions and mental states (Tomasello *et al.*, 2005; Frith and Frith, 2007). The neural underpinnings of intention understanding have been vastly investigated in non-verbal social interactions (Amodio and Frith, 2006; Frith and Frith, 2006; Van Overwalle, 2009; Schurz *et al.*, 2014); yet, whether these findings generalize to the most frequent form of human interactive behaviour—verbal communication and speech—is currently poorly understood. This is surprising given that the main purpose of language use is to convey communicative intentions to take effect on listeners' reactions (Bühler, 1934; and later Grice, 1957; Austin, 1962; Searle, 1969). Remarkably, interlocutors naturally grasp others' intentions even if they are not literally stated (Holtgraves, 2005; Bašnáková *et al.*, 2014)—take an 'excuse me' meant to express criticism. They do so by using contextual (e.g. Clark and Carlson, 1981) or extralinguistic cues (e.g. Frith, 2009), among them the speaker's prosody—her vocal tone (Hellbernd and Sammler, 2016). Here, we investigate whether established mechanisms of social intention recognition apply to the spoken communicative domain and how they interface with the neural auditory prosodic system to transform vocal tone into a coherent representation of the speaker's intent.

Received: 22 August 2017; **Revised:** 6 May 2018; **Accepted:** 13 May 2018

Previous behavioural investigations have shown that interlocutors use *speech prosody*—that is, the rhythmic melodic aspects of speech—as one salient indicator for communicative intent: From early on, both young infants (Dore, 1975; Prieto *et al.*, 2012) and mothers (Fernald, 1989; Papoušek *et al.*, 1990) convey simple intentions, like requests or complaints, through distinct prosodic contours. Similarly, adult conversation contains distinct prosodic markers for various interpersonal attitudes (Aubergé *et al.*, 2003; Jiang and Pell, 2015, 2016) and intentions such as criticism, wish or suggestion (Hellbernd and Sammler, 2016). What remains unresolved so far is how prosody translates into social meaning at the *neurocognitive* level (Jiang *et al.*, 2017; Lavan *et al.*, 2017). On the one hand, distinct prosodic signatures make it plausible to assume *auditory prosodic categorization* processes that may link conventionalized acoustic feature configurations to communicative meaning. On the other hand, given that prosody conveys information on the speaker's intent, these processes are likely to interface with domain general *sociocognitive systems* (e.g. Van Overwalle, 2009; Schurz *et al.*, 2014). The present study weighs the contribution of auditory prosodic and sociocognitive processes and their interaction by combining a novel audio-morphing paradigm with functional magnetic resonance imaging (fMRI).

On the *auditory prosodic* side, prosodic information is known to cascade through several processing levels in lateral frontotemporal and subcortical areas (for reviews, see Baum and Pell, 1999; Belyk and Brown, 2014; Frühholz *et al.*, 2016; Paulmann, 2016). More generally, acoustic signals are thought to attain meaning by passing through consecutive stages along the ventral auditory stream (Rauschecker and Scott, 2009; Weiller *et al.*, 2011; Bajada *et al.*, 2015). Along the temporal lobe, auditory information undergoes gradual abstraction (Patterson *et al.*, 2002; Obleser and Eisner, 2009), from extraction of low-level acoustic features in Heschl's gyrus to representations of abstract sound categories, including speech (Leaver and Rauschecker, 2010; DeWitt and Rauschecker, 2012; Norman-Haignere *et al.*, 2015) and non-speech sounds (Belin *et al.*, 2004; von Kriegstein and Giraud, 2004), in more anterior temporal regions. Recent evidence for right ventral pathways for prosody perception (Schirmer and Kotz, 2006; Frühholz *et al.*, 2015; Sammler *et al.*, 2015) raised proposals of a similar progressive abstraction of prosodic information. Yet, high-level prosodic percept formation, that is, abstraction of prosodic *content* that dissociates from low-level acoustic *form* remains to be shown (Bestelmeyer *et al.*, 2014). To explore abstract encoding of prosodic categories in the ventral auditory stream, the present study employed a categorization task along prosodic continua while carefully controlling for low-level acoustic features.

On the *sociocognitive* side, intention understanding involves inferential processes based on a 'theory of mind' (ToM), also referred to as 'mentalizing' (Schurz *et al.*, 2014). ToM is the ability to ascribe mental states to others and typically activates a brain network including the temporoparietal junction (TPJ) and medial prefrontal cortex (mPFC; Amodio and Frith, 2006; Ciaramidaro *et al.*, 2007; Van Overwalle, 2009). Notably, not only the perception of others as *intentional agents* showed ToM activations (Gallagher *et al.*, 2002; Sripada *et al.*, 2009; Suzuki *et al.*, 2011), for example, when inferring others' (private) action goals (de Lange *et al.*, 2008; Walter *et al.*, 2009; Sebastian *et al.*, 2012). Particularly, the recognition of *social communicative intentions* triggered ToM processes in TPJ and mPFC (Ciaramidaro *et al.*, 2007; Walter *et al.*, 2009; Ciaramidaro *et al.*, 2014) as tested with visual pictorial (Walter *et al.*, 2004; Canessa *et al.*, 2012) or written verbal material (Bohrn *et al.*, 2012; Spotorno *et al.*, 2012; Bašnáková *et al.*, 2014; Egorova *et al.*, 2014). The present study seeks to extend these findings to the vocal domain by limiting social cues to prosodic information only.

One *link* between auditory prosodic and sociocognitive processes during understanding of prosodic intentions might lie in the sound's social relevance. It is well established that socially relevant compared to less relevant stimuli (e.g. emotional vs. neutral prosody) are processed differently, specifically in the amygdala (Sander *et al.*, 2003; Ethofer *et al.*, 2009; Wiethoff *et al.*, 2009; Frühholz and Grandjean, 2013). The amygdala is a highly connected hub (Roy *et al.*, 2009) with strong connections both to sensory (Vuilleumier and Pourtois, 2007; Frühholz *et al.*, 2014) and to mentalizing areas (Li *et al.*, 2014). It may, hence, be in the ideal position to aid and bind sensory perception (Pourtois *et al.*, 2013) and evaluation of socially relevant signals (Adolphs, 2010). To explore functional links between the amygdala on the one hand and auditory prosodic and socio-inferential systems on the other hand, we computed psychophysiological interactions (PPI) with seeds in bilateral amygdala.

In sum, the present study investigated the contribution of auditory prosodic categorization, social inference and their interaction during recognition of prosodically coded communicative intentions. To do so, we employed a novel paradigm that compared *typical* prosodic exemplars of three intentional categories (criticism, doubt and suggestion) with socially *ambiguous* prosodic expressions that had been generated by means of pairwise audio-morphing between the typical prosodies (STRAIGHT; Kawahara, 2006) (Figure 1A). This approach capitalizes on two phenomena: First, audio-morphing quasi-neutralizes intentional meaning around the ambiguous centres of the continua. Second, ambiguous and typical tokens fully match in their average acoustic feature compositions, as they are derived from identical originals. Consequently, brain areas more strongly involved during behavioural classification of *typical* compared to *ambiguous* intention expressions should reflect (i) perception of abstract auditory prosodic categories beyond their acoustic form as well as (ii) social–inferential comprehension of communicative intentions that may (iii) be functionally linked via the amygdala.

Methods

Participants

Twenty-two native German speakers (11 females; mean age 27.35 ± 4.39 SD; 21 of them right-handed according to the Edinburgh Handedness Inventory by Oldfield, 1971; mean laterality quotient 83.00 ± 36.35 *SD*) with normal hearing and no reported history of neurological or psychiatric disorder took part in the fMRI study. Eight more participants had been invited but were excluded because they had not been able to perform the task in a training session ($n = 4$), had excessive motion artefacts ($n = 1$) or fell asleep during scanning ($n = 3$). Prior to the experiment, participants gave written informed consent according to the procedures approved by the Ethics Committee of the University of Leipzig (404/14-ff). They were paid 8€ per hour for their participation.

Stimuli

Speech stimuli were created based on a subset of the stimuli used by Hellbernd and Sammler (2016). This subset comprised the German words 'Bier' (beer) and 'Bar' (bar) uttered by two trained female German speakers (voice coaches) in three intonations expressing the intentions criticism, doubt and suggestion (44.1 kHz, 16 bit, mono; for further details on stimulus

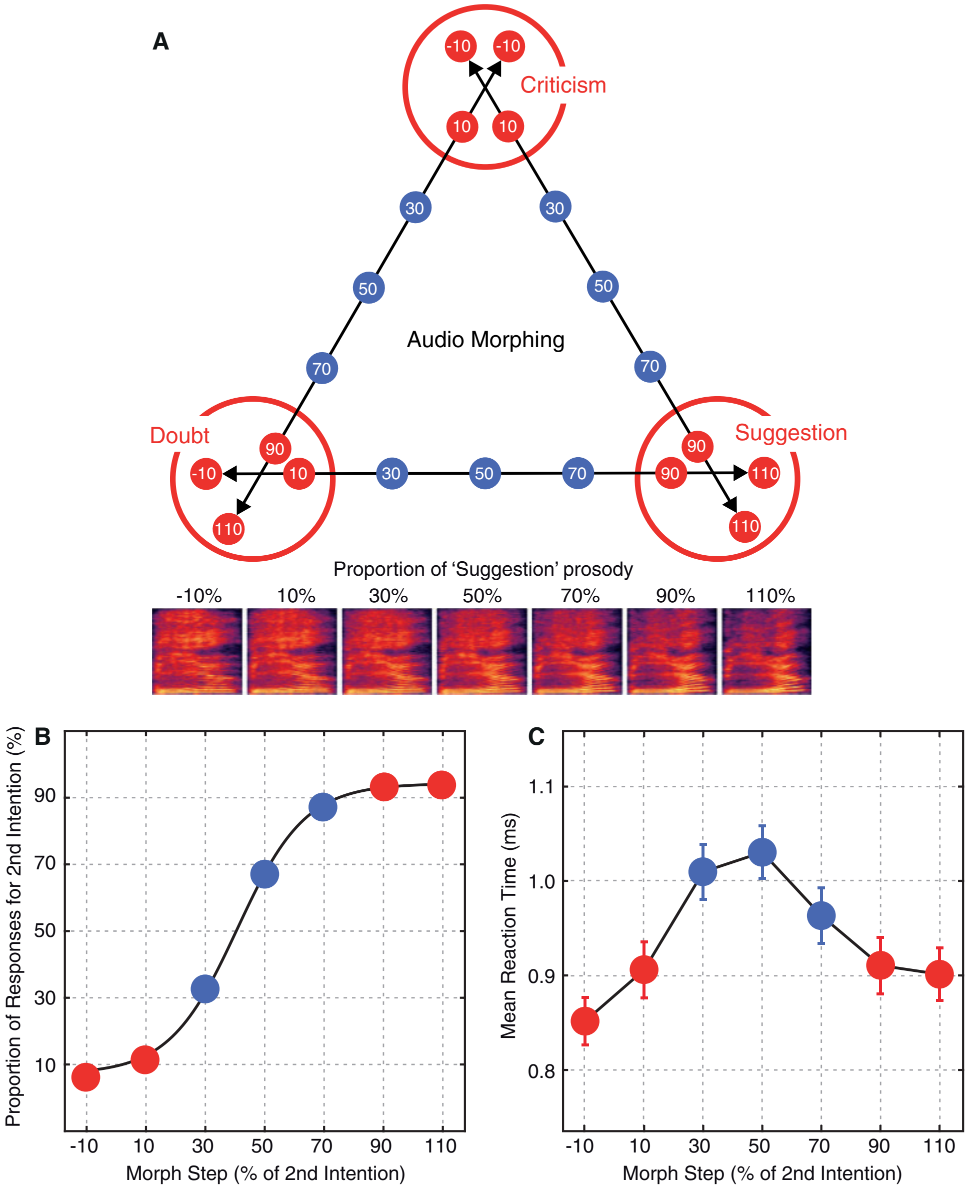

Fig. 1. (A) Experimental stimulus creation through audio-morphing. Continua between doubt and suggestion, criticism and doubt and criticism and suggestion were created with STRAIGHT (Kawahara, 2006). Acoustic features were mixed in consecutive 20% steps. Red dots indicate CLEAR stimuli with only ±10% physical distance from original sounds. Blue dots show AMBIGUOUS stimuli. Spectrograms in the bottom panel exemplify the acoustic transition in the doubt–suggestion continuum. (B & C) Behavioural results. Prosodies with CLEAR intentions (red) were classified more consistently (B) and faster (C) than AMBIGUOUS prosodies (blue).

recording see Hellbernd and Sammler, 2016). These intentions are all characterized by rising pitch contours, avoiding discriminability purely based on pitch direction.

These 12 original stimuli (2 words × 2 speakers × 3 intentions) were downsampled to 16 kHz and subjected to audio-morphing to obtain seven-step continua in which prosody gradually changed in 20% steps from one intention to another (see Figure 1A). Morphing was done with STRAIGHT (Kawahara, 2006), separately for each speaker and each word, resulting in a total of 12 continua: 4 from criticism to doubt, 4 from doubt to suggestion, and 4 from criticism to suggestion. STRAIGHT decomposes the audio stimuli into five parameters: fundamental frequency (f0), formant frequencies, duration, spectrotemporal density and aperiodicity. Temporal anchor points for morphing were set to the onsets and offsets of phonation. Spectrotemporal anchors were set to the first to fourth formants at onsets and offsets of formant transition and intention-specific characteristics. Thereafter, stimuli were re-synthesized based on interpolation of the anchor templates (linearly for duration; logarithmically for the remaining four parameters) of two intentions in ratios of −10%/110%, 10%/90%, 30%/70%, 50%/50%, 70%/30%, 90%/10% and 110%/−10% (see Supplementary Material for one example of each continuum).

Importantly, the outer tokens of each continuum (red in Figure 1A) were acoustically very similar to the original stimuli (±10% physical distance from originals). They were, hence, clear representatives of the respective prosodic category. In turn, the inner tokens with more balanced mix ratios (blue in Figure 1A) were acoustically ambiguous and fell between two prosodic categories. Listeners reported no specific meaning emerging in these stimuli. Average stimulus duration was 475 ms (±45 ms *SD*). Further acoustic properties of the stimuli are depicted in Table 1.

Table 1. Acoustic properties and affect ratings of the morph steps in the seven-step prosody continua

Cont.	Acoustic features								Affect ratings	
	Number of voiced frames	Mean f0 (Hz)	*SD* f0 (Hz)	Mean HNR (dB)	Mean Intensity (dB)	Offset–onset f0 (Hz)	Spectral centre of gravity (Hz)	*SD* Spectrum (Hz)	Valence ratings[a]	Arousal ratings[b]
C	439.0 ± 29.1	279.1 ± 29.6	69.8 ± 17.2	13.1 ± 4.3	66.5 ± 3.1	133.7 ± 58.5	778.3 ± 385.1	577.2 ± 146.4	3.3 ± 1.8	6.8 ± 1.5
	443.5 ± 30.4	266.4 ± 25.5	59.0 ± 19.5	14.4 ± 4.4	67.1 ± 2.4	102.6 ± 49.8	728.4 ± 366.9	558.8 ± 141.1	3.7 ± 1.5	5.6 ± 1.9
	448.0 ± 25.9	255.4 ± 21.5	51.3 ± 20.5	15.2 ± 4.4	67.7 ± 1.9	112.7 ± 54.0	683.8 ± 337.9	561.8 ± 165.8	4.3 ± 1.3	4.8 ± 1.7
↕	456.0 ± 18.3	245.6 ± 17.1	46.0 ± 18.7	15.4 ± 3.9	68.1 ± 1.8	95.0 ± 37.1	631.0 ± 286.4	565.0 ± 163.1	4.4 ± 1.4	4.1 ± 1.7
	461.8 ± 18.1	236.2 ± 11.9	44.0 ± 13.1	15.2 ± 4.0	68.2 ± 1.8	95.6 ± 20.4	594.6 ± 247.6	572.5 ± 176.1	4.3 ± 1.2	3.7 ± 1.6
	467.8 ± 17.9	228.1 ± 9.0	43.9 ± 5.9	14.8 ± 3.8	68.4 ± 1.7	74.1 ± 21.3	563.7 ± 213.6	583.1 ± 196.7	4.1 ± 1.3	3.4 ± 1.5
D	474.3 ± 22.8	220.8 ± 8.8	45.0 ± 4.7	14.2 ± 3.4	68.5 ± 1.4	74.1 ± 16.3	556.0 ± 198.6	640.8 ± 286.7	4.0 ± 1.4	3.1 ± 1.4
Avg	455.8 ± 23.2	247.4 ± 17.6	51.3 ± 14.2	14.6 ± 4.0	67.8 ± 2.0	98.2 ± 36.8	648.0 ± 29.9	579.9 ± 182.3	4.0 ± 1.4	4.5 ± 1.6
C	360.0 ± 21.9	279.2 ± 30.7	67.5 ± 18.2	12.7 ± 4.2	66.5 ± 2.7	122.9 ± 68.7	783.4 ± 388.7	591.9 ± 168.1	3.4 ± 1.6	6.7 ± 1.6
	360.0 ± 18.2	270.4 ± 26.8	61.0 ± 19.3	13.7 ± 4.2	66.8 ± 2.6	135.5 ± 58.0	726.4 ± 369.6	546.5 ± 130.1	4.0 ± 1.8	6.0 ± 1.7
	362.5 ± 16.9	263.6 ± 23.7	58.9 ± 20.4	14.8 ± 4.3	67.0 ± 2.5	154.7 ± 48.1	687.7 ± 357.8	502.0 ± 96.3	4.5 ± 1.2	4.9 ± 1.8
↕	364.3 ± 18.2	257.7 ± 20.7	60.5 ± 20.5	15.7 ± 3.9	67.2 ± 2.4	175.0 ± 45.1	665.4 ± 353.0	472.5 ± 66.0	5.3 ± 1.4	4.7 ± 1.7
	364.3 ± 18.4	252.2 ± 17.3	65.2 ± 20.0	16.6 ± 3.8	67.4 ± 2.3	198.5 ± 35.1	647.8 ± 341.4	452.7 ± 33.8	5.8 ± 1.3	4.2 ± 1.9
	366.0 ± 18.6	248.3 ± 14.4	72.9 ± 20.5	17.0 ± 3.9	67.5 ± 2.2	221.0 ± 34.9	641.7 ± 337.9	439.2 ± 13.7	6.4 ± 1.0	4.4 ± 2.0
S	365.3 ± 16.6	245.8 ± 10.9	82.3 ± 23.0	16.1 ± 3.8	67.6 ± 2.1	236.8 ± 52.6	638.4 ± 331.4	438.4 ± 30.2	6.7 ± 1.1	4.2 ± 1.7
Avg	363.2 ± 18.4	259.6 ± 20.6	66.9 ± 20.3	15.2 ± 4.0	67.1 ± 2.0	177.8 ± 48.9	684.4 ± 354.3	491.9 ± 76.9	5.1 ± 1.4	5.0 ± 1.8
D	397.8 ± 38.9	221.7 ± 9.3	43.1 ± 1.1	13.2 ± 3.2	68.5 ± 1.6	83.1 ± 9.6	570.7 ± 210.4	629.2 ± 239.9	4.3 ± 1.4	3.2 ± 1.3
	393.8 ± 39.8	225.1 ± 8.6	45.8 ± 5.2	14.7 ± 3.7	68.6 ± 1.7	100.8 ± 12.2	555.3 ± 224.1	562.3 ± 174.8	4.7 ± 1.2	3.1 ± 1.4
	398.8 ± 32.8	229.7 ± 9.2	51.4 ± 7.9	15.8 ± 3.9	68.5 ± 1.7	126.3 ± 43.6	555.8 ± 246.1	514.6 ± 122.5	4.9 ± 1.1	3.2 ± 1.5
↕	397.8 ± 35.1	233.9 ± 9.4	57.0 ± 11.0	16.7 ± 4.0	68.4 ± 1.8	154.1 ± 26.7	570.8 ± 271.6	480.4 ± 80.4	5.4 ± 0.9	3.3 ± 1.7
	395.5 ± 35.9	238.1 ± 10.4	63.6 ± 14.1	17.4 ± 3.7	68.2 ± 1.9	184.4 ± 10.2	594.9 ± 297.5	456.2 ± 38.1	5.8 ± 1.2	3.6 ± 1.6
	394.0 ± 36.8	243.1 ± 11.8	71.5 ± 17.6	17.7 ± 3.7	67.9 ± 2.2	218.3 ± 42.2	622.6 ± 320.6	442.1 ± 8.0	6.3 ± 1.4	4.0 ± 1.7
S	395.0 ± 35.9	249.7 ± 12.1	82.0 ± 22.7	16.3 ± 4.0	67.4 ± 2.8	252.1 ± 59.9	660.0 ± 349.0	439.6 ± 38.4	6.6 ± 1.2	4.5 ± 1.8
Avg	396.1 ± 36. 5	234.5 ± 10.1	59.2 ± 11.4	16.0 ± 3.8	68.2 ± 2.0	159.9 ± 30.6	590.0 ± 272.9	503.5 ± 100.3	5.4 ± 1.2	3.6 ± 1.6

Note. Values depict mean ± *SD*. *SD* = standard deviation, f0 = fundamental frequency, HNR = harmonics-to-noise ratio. All values were extracted using PRAAT 5.3.01 (http://www.praat.org).

[a] Valence ratings: 1 = negative, 9 = positive;

[b] Arousal ratings: 1 = calm; 9 = aroused. Cont.: continuum; C: criticism; D: doubt; S: suggestion; Avg: average.

It is of note that clear and ambiguous stimuli, that is, morph steps {1, 2, 6, 7} vs. {3, 4, 5}, did not differ in their averaged acoustic features: mean fundamental frequency (f0), f0 variance (*SD* f0), f0 range (offset f0–onset f0), harmonics-to-noise ratio, centre of gravity and spectral variance (*SD* of spectrum) (independent-samples *t*-tests, all $Ps > .1$; see Table 2). Consequently, neural activation differences between clear and ambiguous prosodies cannot be explained by pure acoustic differences between stimuli but are more likely related to the communicative content and typicality of the utterances.

Finally, we asked 16 listeners (8 female, mean age 24.1 ± 4 *SD*; none of whom participated in the fMRI experiment) to rate each stimulus' valence and arousal using the nine-step self-assessment manikin (Bradley and Lang, 1994). Clear and ambiguous stimuli differed in terms of arousal in all intention combinations and in terms of valence in criticism–doubt continua (see Table 2). To account for these affective differences, valence and arousal ratings were later included as parametric modulators in the fMRI data analysis (see below).

Procedure

Audio morph stimuli were presented in separate blocks for each continuum – that is, separately for each word and speaker – with randomized block order across participants. Within each block, each morph step was presented eight times according to a type-1 index-1 sequence (Nonyane and Theobald, 2007), that is, each morph step followed all other morph steps with the same probability to control for adaptation effects (see Bestelmeyer *et al.*, 2010). Additionally, each block contained eight silent trials. Each trial lasted 2.5 s within which a response had to be given (except for silent trials). Participants classified the stimuli in a 2-alternative forced choice task with the two intentions of the current block continuum as possible responses. The corresponding response options were presented on screen throughout the block. Participants responded by button press with the index and middle finger of the right hand. Key assignment was balanced across blocks and participants. Blocks were separated by 30-s breaks. Total experiment duration (including 12 blocks) was ∼40 min. All participants took part in a screening session with all stimuli on a separate day before the fMRI experiment to ensure that they were able to do the task and a short reminder session (1/4 of the stimuli) on the day of scanning prior to entering the scanner.

After scanning, participants filled out a debriefing questionnaire assessing on nine-step rating scales how strongly they relied on auditory prosodic or sociocognitive strategies to classify the speaker's prosodic intentions.

Data acquisition

fMRI images were acquired with a 3T Siemens Magnetom Prisma scanner (Siemens AG, Erlangen, Germany) at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. During the main experiment, 1155 event-related T_2*-weighted scans were obtained with a single-shot, echo-planar

Table 2. Statistical comparisons of acoustic stimulus features and affect ratings between clear and ambiguous stimuli

Cont.	Acoustic features[a]																Affect ratings[b]			
	Voiced frames		Mean f0		SD f0		Mean HNR		Mean Intensity		Offset-onset f0		Spectral CoG		SD Spectrum		Valence ratings		Arousal ratings	
	t(26)	*P*	*t*(26)	*P*	*t*(26)	*P*	*t*(26)	*P*	*t*(26)	*P*	*t*(26)	*P*	*t*(26)	*P*	*t*(26)	*P*	*t*(26)	*P*	*t*(26)	*P*
C – D	0.10	0.92	0.30	0.77	1.17	0.25	−0.80	0.43	−0.55	0.59	−0.32	0.75	0.19	0.85	0.37	0.71	−5.14	0.00	3.78	0.00
C – S	−0.14	0.89	0.37	0.71	1.29	0.21	−0.58	0.57	−0.13	0.90	0.14	0.89	0.25	0.81	0.80	0.43	−0.71	0.49	10.69	0.00
D – S	−0.18	0.86	0.21	0.83	0.53	0.60	−0.82	0.42	−0.41	0.68	0.00	0.71	0.30	0.77	0.75	0.46	1.98	0.07	2.42	0.03

Note. Cont.: continuum; C – D: criticism—doubt; C – S: criticism–suggestion; D – S: doubt–suggestion; HNR: harmonics-to-noise ratio; CoG: center of gravity.

[a] Independent-samples *t*-tests.

[b] Paired-samples *t*-tests.

gradient-echo (EPI) sequence (repetition time (TR) = 2000 ms, echo time (TE) = 26 ms). Forty axial slices (3 mm isotropic voxel size, 10% inter-slice gap) were collected with a 20-channel head coil. The field of view (FoV) was 192 × 192 mm^2 with an in-plane resolution of 64 × 64 pixels and a flip angle of 90°.

Auditory stimuli were presented via MR-compatible headphones (MR confon GmbH, Magdeburg, Germany). Before scanning, we made sure that participants heard the sounds clearly in the scanner.

High-resolution T_1-weighted images (1 mm isotropic voxel size) for anatomical coregistration acquired with a magnetization-prepared rapid acquisition gradient echo sequence were available for all participants in the brain database of the Max Planck Institute for Human Cognitive and Brain Sciences.

Functional MR data analysis

Functional MR images were preprocessed and analysed using SPM8 (Wellcome Trust Centre for Neuroimaging, London, United Kingdom) in Matlab. During preprocessing, the first four volumes were discarded to account for T_1-equilibration effects. To correct for distortion and motion, functional images were realigned and unwarped. Thereafter, images were co-registered with the anatomical image and normalized into Montréal Neurological Institute (MNI) space (without resampling). Finally, smoothing was applied using a Gaussian kernel of 8 mm full width at half maximum (FWHM).

At the first level, a general linear model (Friston *et al.*, 1994) was applied to the data of every participant. Twenty-one onset regressors – one for each morph step per continuum, that is, criticism–doubt, doubt–suggestion, criticism–suggestion, collapsed across words and speakers – were convolved with a canonical hemodynamic response function. Events were time-locked to stimulus onset, and a high-pass filter with a cut-off frequency of 1/128 Hz was applied. Z-transformed reaction times of each participant as well as mean arousal and valence ratings for each stimulus (obtained from 16 independent raters; see Table 1) were included as parametric modulators to account for task difficulty and affective connotations of the stimuli. Additionally, six motion parameters were entered as covariates of no interest to control for subtle head movements. The linear contrasts between all outer stimuli of the morphed continua with CLEAR prosodies (Figure 1, red) and all inner morphs with AMBIGUOUS prosodies (Figure 1, blue) were calculated for each participant (i.e. CLEAR > AMBIGUOUS and AMBIGUOUS > CLEAR) and submitted to one-sample *t*-tests during random effects group analysis. As an aside, results obtained with this analysis did not differ from positive and negative quadratic effects obtained with a parametric approach using three onset regressors (one for each continuum), each followed by four orthogonalized parametric modulators: (1) z-transformed reaction times, (2) mean arousal, (3) mean valence ratings, and (4) morph step (for a similar approach, see Bestelmeyer *et al.*, 2014). Statistical threshold was set to a conservative $P < 0.0001$ at voxel level, followed by cluster-level family-wise error (FWE) correction of $P < 0.05$.

Psychophysiological interaction analysis

A psychophysiological interaction (PPI) analysis was conducted to identify task-dependent changes in amygdala connectivity. In order to define voxels of interest (VOIs), bilateral amygdala regions were defined based on the anatomical Harvard-Oxford atlas in FSL (Jenkinson *et al.*, 2012) for all voxels with likelihood > 20% for that region. Within these regions, all voxels with single-subject activations below a threshold of $P = 0.05$ were included in the analysis. The first eigenvariate of the fMRI signal changes in the VOIs was extracted, and their mean time course was multiplied by a regressor with information about the experimental conditions (CLEAR vs. AMBIGUOUS). This interaction term of source signal and experimental treatment was the first regressor in the PPI analysis. Additionally, the mean deconvolved source signal of the VOI and the task regressor were included in the model as covariates of no interest. Participants without activations in the amygdala at single-subject level were excluded from the PPI analyses, resulting in a sample size of $n = 20$ for left amygdala and $n = 17$ for the right amygdala seed. Statistical threshold was $P < 0.0001$ at voxel level, followed by cluster-level FWE correction of $P < 0.05$.

Results

Behavioural data

Figure 1B demonstrates that participants reliably recognized the intentions expressed by the CLEAR stimuli (i.e. values lower than 15% and higher than 90%), whereas AMBIGUOUS morph steps were classified less consistently. Furthermore, reaction times were significantly faster for CLEAR (mean ± SD: 892 ± 131 ms) than for AMBIGUOUS stimuli (1001 ± 135 ms) as tested with a paired-samples *t*-test ($t(21) = -9.817$, $P < 0.0001$; Figure 1C).

fMRI data

CLEAR compared to AMBIGUOUS prosodic expressions induced stronger activations in areas known for auditory prosodic as

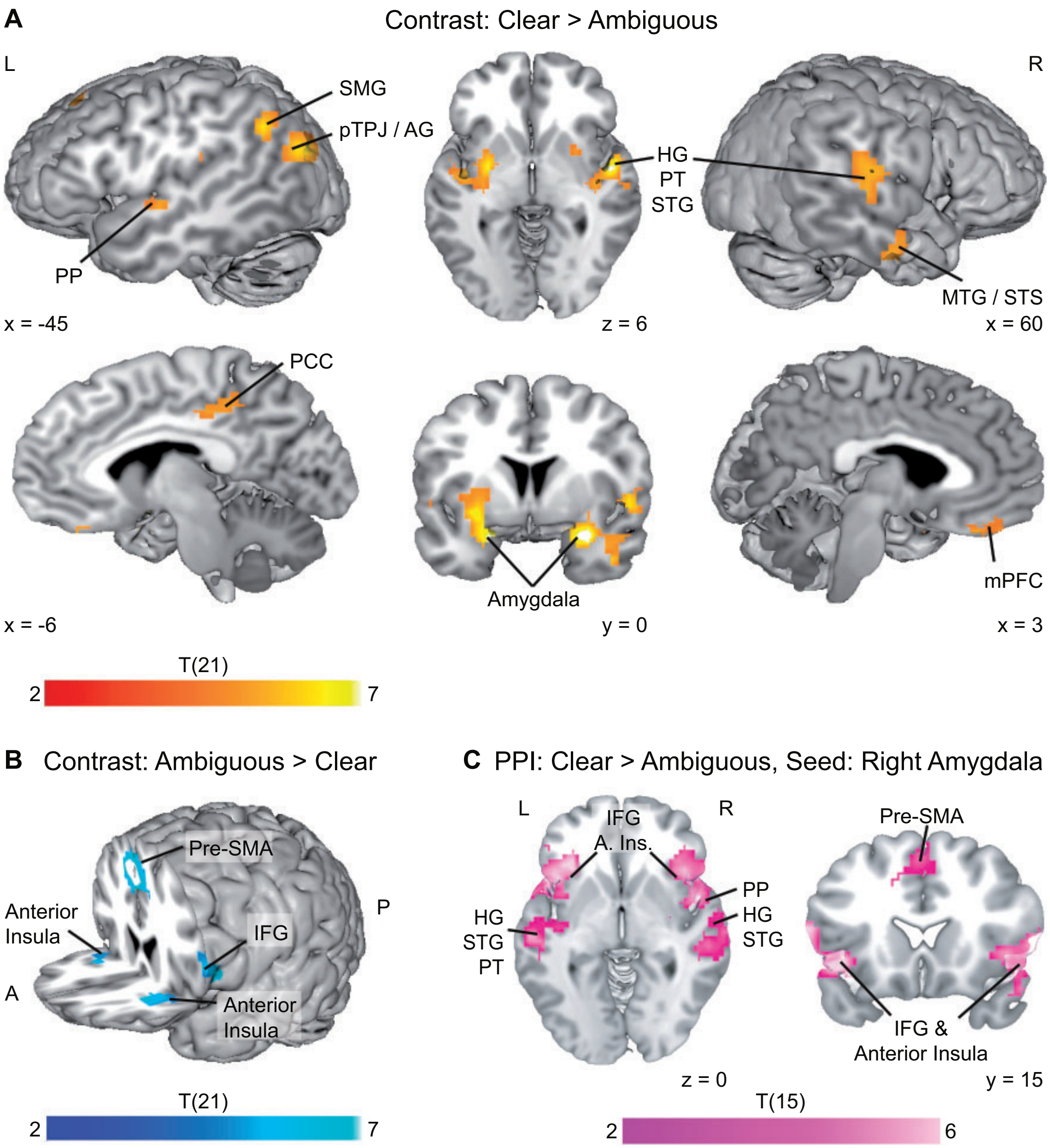

Fig. 2. Activation maps and PPI analysis. (A) CLEAR > AMBIGUOUS prosodies activated auditory prosodic as well as sociocognitive brain regions. (B) AMBIGUOUS > CLEAR prosodies activated cingulo-opercular areas in both hemispheres. Threshold: voxel $P < 0.0001$, cluster $P < 0.05$ FWE-corrected. (C) Amygdala was functionally connected with auditory prosodic as well as cingulo–opercular regions, more strongly during clear than ambiguous prosodies. Results are displayed for the right amygdala seed only (for the similar results of the left amygdala seed, see Table 4). Threshold: voxel $P < 0.001$, cluster $P < 0.05$ FWE-corrected. SMG: supramarginal gyrus; pTPJ: posterior temporoparietal junction; AG: angular gyrus; PP: planum polare; STG/STS: superior temporal gyrus/sulcus; MTG: middle temporal gyrus; HG: Heschl's gyrus; PT: planum temporale; PCC: posterior cingulate cortex; mPFC: medial prefrontal cortex; pre-SMA: pre supplementary motor area; IFG: inferior frontal gyrus; FWE: family-wise error.

well as sociocognitive brain regions (Figure 2A; Table 3). Auditory prosodic activations comprised right Heschl's gyrus (HG), planum temporale (PT) and posterior superior temporal gyrus (STG) as well as left planum polare (PP). Sociocognitive areas included bilateral amygdala and hippocampus, ToM regions including bilateral mPFC and posterior cingulate cortex (PCC), right anterior middle temporal gyrus (MTG) and superior temporal sulcus (STS) as well as left posterior TPJ/angular gyrus (pTPJ/AG), supramarginal gyrus (SMG). Additionally, left middle frontal gyrus (MFG) and left basal ganglia (putamen and pallidum) were active.

Notably, activation of mPFC (at peak MNI coordinate [−6, 35, −20]) positively correlated with participants' focus on the speaker's perspective as assessed in the debriefing questionnaire ('Did you focus most on the acoustic sound of the stimulus (low score) or the speaker's perspective (high score)?'; $r(20) = 0.43$, $P = 0.044$).

In the opposite contrast, AMBIGUOUS compared to CLEAR prosodies evoked stronger activations in presupplementary

Table 3. Functional activations for the contrasts Clear > Ambiguous and Ambiguous > Clear

Brain region	Hem.	BA	k	x	y	z	Z-value
Clear > ambiguous							
Amygdala[a]	**R**	**–**	**774**	**30**	**2**	**−14**	**5.33**
Hippocampus[a]	R	–		33	−16	−17	4.68
Heschl's gyrus[a]	R	41/42		48	−7	1	5.17
Planum temporale	R	22		57	−22	7	4.40
Posterior superior temporal gyrus	R	22		66	−34	16	4.19
Anterior superior temporal sulcus	R	22/21		48	−4	−17	4.52
Anterior middle temporal gyrus	R	21		57	−7	−20	4.47
Central operculum	R	48		54	−13	13	4.58
Amygdala[a]	**L**	**–**	**362**	**−21**	**−1**	**−17**	**5.01**
Putamen[a]	L	–		−27	−1	−2	4.87
Hippocampus	L	–		−27	−19	−17	3.88
Pallidum	L	–		−21	−4	7	4.15
Planum polare	L	22		−42	−13	−5	4.61
Central operculum	L	48		−51	−4	7	3.79
Posterior temporoparietal junction[a]	**L**	**39**	**281**	**−42**	**−70**	**22**	**4.89**
Angular gyrus[a]	L	40		−54	−58	34	4.87
Supramarginal gyrus	L	40		−60	−40	40	4.52
Parietal operculum[a]	**L**	**48**	**35**	**−54**	**−28**	**19**	**4.68**
Posterior cingulate cortex	**L**	**–**	**88**	**−12**	**−34**	**40**	**4.58**
Posterior cingulate cortex	**R**	**–**	**27**	**12**	**−22**	**43**	**4.04**
Medial prefrontal cortex	**L**	**11**	**50**	**−6**	**35**	**−20**	**4.36**
Paracingulate gyrus	L	11		−3	38	−8	3.75
Medial prefrontal cortex	R	11		6	41	−17	4.02
Middle frontal gyrus	**L**	**9**	**48**	**−27**	**26**	**46**	**4.34**
Ambiguous > Clear							
Presupplementary motor area[a]	**R**	**6/32**	**360**	**6**	**14**	**49**	**5.57**
Paracingulate gyrus[a]	R	32		3	26	40	5.47
Presupplementary motor area[a]	L	6/32		−3	11	55	5.26
Frontal orbital cortex / anterior insula	**L**	**47/–**	**59**	**−30**	**32**	**−2**	**4.83**
Inferior frontal gyrus (p. op.)	**L**	**44**	**57**	**−54**	**14**	**13**	**4.75**
Frontal orbital cortex / anterior insula	**R**	**47/–**	**54**	**30**	**29**	**−2**	**4.44**
Anterior Insula	R	–		36	23	1	3.98

Note. BA: Brodmann area; L: left hemisphere; R: right hemisphere; k: cluster extent (number of voxels); p. op.: pars opercularis; p. tri.: pars triangularis. Coordinates indicate cluster peaks in MNI-space. Main peaks are in bold. P-voxel < 0.0001, P-cluster < 0.05 FWE-corrected.

[a]Peaks that are significant at P-voxel < 0.05 FWE-corrected.

motor area (pre-SMA), bilateral frontal orbital cortex and anterior insula and left inferior frontal gyrus (IFG) (Figure 2B; Table 3).

PPI analysis

The PPI analysis revealed significant connectivity changes of left and right amygdala as a function of CLEAR or AMBIGUOUS prosodies (see *Methods*). Connectivity was enhanced between bilateral amygdala on the one hand and bilateral anterior insula, frontal orbital cortex and IFG on the other hand during recognition of CLEAR intentions in prosody. Additionally, right amygdala showed higher connectivity during CLEAR stimuli with bilateral auditory regions including HG, left PT and right PP, bilateral intraparietal sulcus and pre-SMA as well as left dorsal (pre)motor areas (Figure 2C; Table 4).

Discussion

A speaker's vocal tone – her prosody – is an important carrier of social information. Interpersonal attitudes, private beliefs and communicative intentions resonate in a speaker's voice and shape the tenor and success of social interactions. However, to date there are no explanatory accounts of the neurocognitive mechanisms that allow listeners to 'read' speakers' intents from vocal prosodic patterns in speech. The present fMRI study identified two sets of distributed brain regions typically associated with (i) auditory prosodic categorization and (ii) social–inferential conceptualization that were both involved when listeners recognized *clearly* expressed intentions in a speaker's voice. *Ambiguous* expressions, in turn, involved (iii) cingulo-opercular regions known to assist categorization in case of conflicting cues. The amygdala took a central position and exhibited flexible functional connections with auditory and decision-making areas that were contingent on the typicality of the prosodic expressions. The combined findings suggest complementary mechanisms via which prosody gains communicative meaning (i) based on abstract encoding of acoustic profiles that link with social concepts along the ventral auditory stream and are recurrently tagged by the amygdala for their social relevance; (ii) based on listeners' propensity to infer the speaker's inner mental state drawing on mentalizing areas and (iii) based on controlled decision-making processes to resolve the meaning of equivocal expressions. Altogether, this work sheds light on the neural complexity of intentional vocal prosodic signalling in speech and illustrates its anchoring at the interface between auditory and social cognition.

Table 4. Results of the PPI analysis

Brain region	Hem.	BA	k	x	y	z	Z-value
Seed: Left Amygdala, N = 20							
Frontal orbital cortex	**L**	**47**	**242**	**−30**	**29**	**−8**	**4.66**
Anterior insula	L	–		−42	14	−5	3.92
Inferior frontal gyrus (p. op.)	L	44		−54	14	7	3.86
Superior temporal gyrus	L	22		−57	−4	−8	3.27
Frontal orbital cortex / anterior insula	**R**	**47/–**	**374**	**42**	**20**	**−5**	**4.35**
Inferior frontal gyrus (p. tri.)	R	45		45	29	7	3.95
Inferior frontal gyrus (p. op.)	R	44		57	20	13	3.54
Central opercular cortex	R	48		48	5	7	3.45
Seed: Right Amygdala, N=16							
Inferior frontal gyrus (p. op./tri.)	**R**	**44/45**	**999**	**57**	**20**	**7**	**4.93**
Anterior insula / frontal orbital cortex	R	–/47		42	20	−8	4.76
Precentral gyrus	R	6		60	5	22	3.53
Planum polare	R	22		51	−1	−5	4.29
Superior temporal gyrus / sulcus	R	22/21		45	−25	−2	3.93
Heschl's gyrus	R	41/42		54	−13	7	3.65
Planum temporale	**L**	**22**	**340**	**−42**	**−37**	**16**	**4.59**
Heschl's gyrus	L	41/42		−54	−10	7	3.67
Central operculum	L	48		−39	−19	16	4.13
Superior temporal gyrus	L	22		−57	−22	1	4.07
Putamen	L	–		−33	−10	4	3.31
Anterior insula	**L**	**–**	**437**	**−42**	**14**	**−2**	**4.53**
Inferior frontal gyrus (p. op.)	L	44		−60	14	10	4.38
Inferior frontal gyrus (p. tri.)	L	45		−51	38	4	3.48
Precentral gyrus	**L**	**6**	**246**	**−15**	**−22**	**76**	**4.51**
Central sulcus	L	4/6		−30	−28	67	3.80
Postcentral gyrus	L	3		−39	−34	61	3.59
Intraparietal sulcus	L	7/40		−27	−46	40	3.56
Intraparietal sulcus	**R**	**7/40**	**152**	**36**	**−43**	**37**	**4.32**
Presupplementary motor area	**L/R**	**6/32**	**102**	**0**	**14**	**52**	**3.61**
Paracingulate gyrus	R	32		3	8	46	3.49

Note. BA: Brodmann area; L: left hemisphere; R: right hemisphere; k: cluster extent (number of voxels).Coordinates indicate cluster peaks in MNI-space. Main peaks are in bold. P-voxel < 0.001, P-cluster < 0.05 FWE-corrected.

Auditory prosodic abstraction

Socially meaningful (compared to ambiguous) prosodic expressions induced activations along the right ventral auditory stream, including early auditory areas in HG, adjacent PT and STG as well as downstream conceptual areas in anterior STS and MTG (Binder and Desai, 2011; Lambon Ralph *et al.*, 2017). In the left hemisphere, activations were limited to a small region in PP. This line-up is in keeping with the relevance of (right) temporal areas for prosody perception (for reviews, see Baum and Pell, 1999; Belyk and Brown, 2014) as well as the recent description of a pathway for prosody along the right temporal lobe (Sammler *et al.*, 2015; see also Schirmer and Kotz, 2006). It is important to note that average physical characteristics did not differ between clear and ambiguous stimuli. This suggests a contribution of these areas to the perception of prosodic *content* – beyond basic acoustic forms – that seems to already emerge at early processing levels. Consistent with this idea, PT and peri-auditory regions of STG are known for the abstract, categorical encoding of auditory objects independent of low-level acoustic features (Griffiths and Warren, 2002; Warren *et al.*, 2005; Kumar *et al.*, 2007) both in speech (Chang *et al.*, 2010; Kilian-Hütten *et al.*, 2011) and non-speech sounds (Giordano *et al.*, 2013; Latinus *et al.*, 2013). Similar constructive processes have been recently reported at levels as early as HG (Formisano *et al.*, 2008; Kilian-Hütten *et al.*, 2011). The present data suggest an extension of these mechanisms to the perception of abstract prosodic categories. Abstract encoding of sound categories was proposed to amount from experience-dependent perceptual biases that accentuate ecologically meaningful rather than physical properties of sensory stimuli (Nelken and Ahissar, 2006), especially when they are behaviourally relevant (for review, see Bizley and Cohen, 2013). A possible source that relays relevance information to auditory cortex and, hence, induces the said perceptual bias is the amygdala (Adolphs, 2010; Kumar *et al.*, 2012; Pourtois *et al.*, 2013), which we found strongly activated during clear (compared to ambiguous) prosodies as will be discussed below.

Apart from this *perceptual* abstraction and categorization of prosodic patterns in auditory areas, recognition of speakers' (clear) intentions further involved the right anterior STS and MTG—areas that are among the targets of the ventral auditory stream (Rauschecker and Scott, 2009) and typically feature as 'representational hub' in semantic cognition (for reviews, see Binder and Desai, 2011; Lambon Ralph *et al.*, 2017). The bilateral anterior temporal lobes (aTLs) are frequently considered as convergence zones, where inputs get detached from their sensory format and gradually turn into coherent concepts that link to meaning (for reviews, Binder *et al.*, 2009; Lambon Ralph, 2013). Auditory lexical signals achieve such *representational* abstraction via dorsal rather than ventral aTL (Skipper *et al.*, 2011; for reviews, see Wong and Gallate, 2012; Olson *et al.*, 2013), compatible with the present activations in anterior STS/MTG. Notably, the emerging concepts are thought to include abstract social knowledge: For example, words denoting social actions (e.g. *to* *embrace*, Lin *et al*., 2015; Lin *et al*., 2017) or social traits (e.g. *loyal*; Zahn *et al*., 2007, 2009; Ross and Olson, 2010; Skipper *et al*., 2011; Binney *et al*., 2016; Pobric *et al*., 2016) evoke robust responses in anterior MTG/STG. Our data suggest that similar social conceptualization processes may be triggered by vocal prosodic cues, beyond the social meaning of words.

Taken together, our results are compatible with a stepwise abstraction of auditory prosodic perceptual patterns towards internal social conceptual representations of the speaker's intention along the temporal lobe. Note that this conversion of sound to meaning is apparently only successful if the acoustic composition of the stimulus is close to the prototypical prosodic signature of the respective communicative intention: Acoustic deviation from the clear representatives of the prosodic category reduced speed and response consistency when labelling the intention. This lends behavioural evidence for the conventionalization of the prosodic forms of intentions that allows listeners to reference the perceived pattern to internalized prosodic prototypes – perhaps through early encoding biases – when inferring the relevant communicative concept (see also Hellbernd and Sammler, 2016).

Social–inferential conceptualization

Understanding (clear) communicative intentions was further accompanied by activations in mPFC, PCC and left pTPJ/AG that together with anterior MTG form the mentalizing network (Amodio and Frith, 2006; Van Overwalle, 2009; Mar, 2011; Schurz *et al*., 2014). On the assumption that this network typically activates when humans think about others' inner mental states (Frith and Frith, 2006; Schurz *et al*., 2014), including their social intentions (Walter *et al*., 2004; Ciaramidaro *et al*., 2007), our data suggest similar mentalizing processes when communicative intentions are inferred from vocal prosodic cues in speech. The observed positive correlation between mPFC activity and reported focus on the speaker's perspective (rather than the sound of the utterance) supports this interpretation. Furthermore, the present involvement of pTPJ/AG and mPFC is in line with their reported activation during inference of unspoken, *implied* meaning in verbal communication (Jang *et al*., 2013; Bašnáková *et al*., 2014; Bögels *et al*., 2015) as in the case of metaphor and irony (Bohrn *et al*., 2012; Spotorno *et al*., 2012) or indirect requests (van Ackeren *et al*., 2012).

Even if our study was not designed to decompose task-specific contributions of the various areas, we note that our task hinges on thoughts about others' *interpersonal* intentions rather than others' *private* beliefs, attitudes or traits. It hence relates tightly to interactions embedded in social situations that participants may have invoked (despite listening to single word prosody) by drawing on past experience. The synthesis of social context and activation of social 'scripts' has been frequently attributed to anterior MTG (Frith and Frith, 2003; Gallagher and Frith, 2003) that we indeed found strongly activated, and that we discussed earlier as a core area of semantic cognition (Binder *et al*., 2009; Binder and Desai, 2011). Note that both views are not mutually exclusive. Rather, the overlap of social and semantic processes raises the interesting hypothesis that anterior MTG constitutes a hub where multimodal percepts and social concepts fuse into meaningful and behaviourally relevant representations (for similar proposals, see Binder and Desai, 2011; Wong and Gallate, 2012). This would illustrate the close relationship between speech communication and mentalizing, mediating between what is said and what is meant.

Functional integration via the amygdala

Apart from these cortical regions, it was the bilateral amygdala with peaks in the superficial nuclei that responded most strongly to clear compared to ambiguous prosodic intentions. Beyond its reputation as 'fear centre' of the brain (LeDoux, 2007) and tracker of visual and auditory emotions (for review, see Ball *et al*., 2009), current accounts stress the amygdala's, and particularly the superficial nuclei's, more general role in processing category typical, socially *relevant* stimuli (Said *et al*., 2009; Adolphs, 2010; Bzdok *et al*., 2013; Stolier and Freeman, 2017; Wang *et al*., 2017), including prosodic expressions (for reviews, see Kotz & Paulmann, 2011; Frühholz *et al*., 2014). More than that, plenty of examples suggest that the amygdala directs resources toward information with subjective impact by perceptually enhancing stimulus features that reveal the social meaning of an expression (for reviews, see Vuilleumier and Pourtois, 2007; Pourtois *et al*., 2013). Dense bidirectional interconnections between amygdala and sensory areas have been proposed as anatomical basis for this 'gain-control' mechanism (Roy *et al*., 2009; Gschwind *et al*., 2012; Frühholz *et al*., 2014). In keeping with these views, we found bilateral amygdala not only to be more strongly activated when prosodic expressions were socially meaningful and, hence, relevant in interpersonal terms. We also saw increased connectivity between right amygdala and bilateral auditory areas (HG, PT, PP and STG) along with stronger auditory activation during clear intentional stimuli, in line with 'gain-control' theories of amygdala function (Pourtois *et al*., 2013). Altogether, we propose that these recurrent loops between amygdala and auditory regions may promote the processing of meaningful prosodic signals and their categorization as behaviourally relevant, intentional sounds.

Some researchers consider the amygdala to be part of the mentalizing system (Mar, 2011), and it is uncontroversial that the amygdala has strong links to various mentalizing areas, for example, mPFC and aTL (Roy *et al*., 2009; Bzdok *et al*., 2013; Li *et al*., 2014). The reasons why these connections were not significantly modulated by the social meaning of our stimuli (see PPI results) remain to be determined. Nevertheless, its special position in the extended mentalizing system as well as in recurrent auditory loops may qualify the amygdala as an ideal interface for sociocognitive and auditory prosodic processes during comprehension of social communicative meaning from vocal cues (see also Leitman *et al*., 2010).

Categorical decision-making

Finally, ambiguous (compared to clear) stimuli evoked stronger activity in pre-SMA, bilateral anterior insula extending into frontal orbital cortex and left IFG (Figure 2C). These cingulo-opercular areas have been frequently implicated in the explicit evaluation of prosody (Belyk and Brown, 2014) and particularly of ambiguous prosodic expressions (Leitman *et al*., 2010; Sammler *et al*., 2015). More generally, these regions respond to a variety of cognitive control processes, including response conflict and interference resolution (Nee, Jonides, and Berman, 2007; Levens and Phelps, 2010). Along these lines, the present activation is likely to reflect controlled processes to arrive at a categorical decision from conflicting prosodic cues. In case of clear prosodic cues, this network showed in turn enhanced connectivity with the amygdala, possibly representing the ideal configuration for naturally smooth categorization and comprehension of socially relevant prosodic signals.

Conclusion

In sum, comprehension of social intentions from a speaker's vocal tone involves the cooperation of distributed (i) auditory prosodic, (ii) sociocognitive and (iii) cingulo-opercular brain areas. Functionally, these sets of regions may be specialized in (i) the abstraction of auditory categories and their mapping against conventionalized prosodic forms, (ii) their association with experience-dependent social semantic concepts and appraisal of the speaker's inner mental state and (iii) the categorical evaluation of (socially ambiguous) stimuli. The amygdala as a social relevance detector flexibly links with auditory perception and categorical decision-making to support the seamless comprehension of socially meaningful information. While future studies should firmly assess these functional hypotheses, our results illustrate the complexity of neural mechanisms through which the brain translates vocal prosodic cues into coherent categorizations of others' intentions as basis for successful interpersonal communication.

Funding

The study was funded by the Otto Hahn Award of the Max Planck Society to DS.

Supplementary data

Supplementary data are available at SCAN online.

Conflict of interest. None declared.

References

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences*, **1191**, 42–61.

Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews: Neuroscience*, **7**(4), 268–77.

Aubergé, V., Audibert, N., Rilliard, A. (2003). Why and how to control the authentic emotional speech corpora. In *Eurospeech-2003*, 185–88.

Austin, J.L. (1962). *How to Do Things with Words*. Cambridge: Harvard University Press.

Bajada, C.J., Lambon Ralph, M.A., Cloutman, L.L. (2015). Transport for language south of the Sylvian fissure: the routes and history of the main tracts and stations in the ventral language network. *Cortex*, **69**, 141–51.

Ball, T., Derix, J., Wentlandt, J., *et al.* (2009). Anatomical specificity of functional amygdala imaging of responses to stimuli with positive and negative emotional valence. *Journal of Neuroscience Methods*, **180**(1), 57–70.

Bašnáková, J., Weber, K., Petersson, K.M., van Berkum, J., Hagoort, P. (2014). Beyond the Language Given: the Neural Correlates of Inferring Speaker Meaning. *Cerebral Cortex*, **24**(10), 2572–8.

Baum, S.R., Pell, M.D. (1999). The neural bases of prosody: insights from lesion studies and neuroimaging. *Aphasiology*, **13**(8), 581–608.

Belin, P., Fecteau, S., Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, **8**(3), 129–35.

Belyk, M., Brown, S. (2014). Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, **9**(9), 1395–403.

Bestelmeyer, P.E.G., Maurage, P., Rouger, J., Latinus, M., Belin, P. (2014). Adaptation to Vocal Expressions Reveals Multistep Perception of Auditory Emotion. *Journal of Neuroscience*, **34**(24), 8098–105.

Bestelmeyer, P.E.G., Rouger, J., DeBruine, L.M., Belin, P. (2010). Auditory adaptation in vocal affect perception. *Cognition*, **117**(2), 217–23.

Binder, J.R., Desai, R.H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, **15**(11), 527–36.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, **19**(12), 2767–96.

Binney, R.J., Hoffman, P., Lambon Ralph, M.A. (2016). Mapping the Multiple graded contributions of the anterior temporal lobe representational hub to abstract and social concepts: evidence from distortion-corrected fMRI. *Cerebral Cortex*, **26**(11), 4227–41.

Bizley, J.K., Cohen, Y.E. (2013). The what, where and how of auditory-object perception. *Nature Publishing Group*, **14**(10), 693–707.

Bögels, S., Barr, D.J., Garrod, S., Kessler, K. (2015). Conversational interaction in the scanner: mentalizing during language processing as revealed by MEG. *Cerebral Cortex*, **25**(9), 3219–34.

Bohrn, I.C., Altmann, U., Jacobs, A.M. (2012). Looking at the brains behind figurative language—a quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, **50**(11), 2669–83.

Bradley, M., Lang, P.J. (1994). Measuring Emotion: the Self-Assessment Manakin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, **25**(1), 49–59.

Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer.

Bzdok, D., Laird, A.R., Zilles, K., Fox, P.T., Eickhoff, S.B. (2013). An investigation of the structural, connectional, and functional subspecialization in the human amygdala. *Human Brain Mapping*, **34**(12), 3247–66.

Canessa, N., Alemanno, F., Riva, F., *et al.* (2012). The neural bases of social intention understanding: the role of interaction goals. *PLoS One*, **7**(7), e42347. doi:10.1371/journal.pone.0042347.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, **13**(11), 1428.

Ciaramidaro, A., Adenzato, M., Enrici, I., *et al.* (2007). The intentional network: how the brain reads varieties of intentions. *Neuropsychologia*, **45**(13), 3105–13

Ciaramidaro, A., Becchio, C., Colle, L., Bara, B.G., Walter, H. (2014). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Social Cognitive and Affective Neuroscience*, **9**(7), 909–16.

Clark, H.H., Carlson, T.B. (1981). Context for Comprehension. In Ling, J., Baddeley, A., editors. *Attention and Performance IX* (pp. 313–330), Hillsdale, NJ: Lawrence Erlbaum Associates.

de Lange, F.P., Spronk, M., Willems, R.M., Toni, I., Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current Biology*, **18**(6), 454–7.

DeWitt, I., Rauschecker, J.P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, **109**(8), E505–14.

Dore, J. (1975). Holophrases, speech acts and language universals. *Journal of Child Language*, **2**(1), 21–40.

Egorova, N., Pulvermüller, F., Shtyrov, Y. (2014). Neural dynamics of speech act comprehension: an MEG study of naming and requesting. *Brain Topography*, **27**(3), 375–92.

Ethofer, T., Kreifelts, B., Wiethoff, S., *et al.* (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, **21**(7), 1255–68.

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: is the Melody the Message?. *Child Development*, **60**(6), 1497–510.

Formisano, E., De Martino, F., Bonte, M., Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, **322**(5903), 970–3.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, **2**(4), 189–210.

Frith, C.D. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **364**(1535), 3453–8.

Frith, C.D., Frith, U. (2006). The neural basis of mentalizing. *Neuron*, **50**(4), 531–4.

Frith, C.D., Frith, U. (2007). Social cognition in humans. *Current Biology*, **17**(16), R724–32.

Frith, U., Frith, C.D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **358**(1431), 459–73.

Frühholz, S., Grandjean, D. (2013). Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex*, **49**(5), 1394–403.

Frühholz, S., Hofstetter, C., Cristinzio, C., *et al.* (2015). Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proceedings of the National Academy of Sciences*, **112**(5), 1583–8.

Frühholz, S., Trost, W., Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology*, **123**, 1–17.

Frühholz, S., Trost, W., Kotz, S.A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, **68**, 96–110.

Gallagher, H.L., Frith, C.D. (2003). Functional imaging of "theory of mind.". *Trends in Cognitive Sciences*, **7**(2), 77–83.

Gallagher, H.L., Jack, A.I., Roepstorff, A., Frith, C.D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, **16**(3 Pt 1), 814–21.

Giordano, B.L., McAdams, S., Zatorre, R.J., Kriegeskorte, N., Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cerebral Cortex*, **23**(9), 2025–37.

Grice, H.P. (1957). Meaning. *The Philosophical Review*, **66**(3), 377–88.

Griffiths, T.D., Warren, J.D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, **25**(7), 348–53.

Gschwind, M., Pourtois, G., Schwartz, S., Ville, D., Van De., Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral Cortex*, 22(7), 1564–76.

Hellbernd, N., Sammler, D. (2016). Prosody conveys speakers' intentions: accoustic cues for speech act perception. *Journal of Memory and Language*, **88**, 70–86.

Holtgraves, T. (2005). The production and perception of implicit performatives. *Journal of Pragmatics*, **37**(12), 2024–43.

Jang, G., Yoon, S.A., Lee, S.E., *et al.* (2013). Everyday conversation requires cognitive inference: neural bases of comprehending implicated meanings in conversations. *NeuroImage*, **81**, 61–72.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M. (2012). FSL. *NeuroImage*, **62**(2), 782–90.

Jiang, X., Pell, M.D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, **66**, 9–34.

Jiang, X., Pell, M.D. (2016). Neural responses towards a speaker's feeling of (un)knowing. *Neuropsychologia*, **81**, 79–93.

Jiang, X., Sanford, R., Pell, M.D. (2017). Neural systems for evaluating speaker (Un)believability. *Human Brain Mapping*, **38**, 3732–49.

Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: perceptually isomorphic decomposition of speech sounds. *Acoustical Science and Technology*, **27**(6), 349–53.

Kilian-Hütten, N., Valente, G., Vroomen, J., *et al.* (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *The Journal of Neuroscience*, **31**(5), 1715–20.

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

Kriegstein, K.V., Giraud, A.-L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, **22**(2), 948–55.

Kumar, S., Stephan, K.E., Warren, J.D., Friston, K.J., Griffiths, T.D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Computational Biology*, **3**(6), e100.

Kumar, S., von Kriegstein, K., Friston, K., Griffiths, T.D. (2012). Features versus feelings: dissociable representations of the acoustic features and valence of aversive sounds. *The Journal of Neuroscience*, **32**(41), 14184–92.

Lambon Ralph, M.A. (2013). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **369**(1634), 20120392.

Lambon Ralph, M.A., Jefferies, E., Patterson, K., Rogers, T.T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, **18**(1), 42–55.

Latinus, M., McAleer, P., Bestelmeyer, P.E.G., Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, **23**(12), 1075–80.

Lavan, N., Rankin, G., Lorking, N., Scott, S., McGettigan, C. (2017). Neural correlates of the affective properties of spontaneous and volitional laughter types. *Neuropsychologia*, **95**, 30–9.

Leaver, A.M., Rauschecker, J.P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *Journal of Neuroscience*, **30**(22), 7604–12.

LeDoux, J. (2007). The amygdala. *Current Biology*, **17**(20), R868–74.

Leitman, D.I., Wolf, D.H., Ragland, J.D., *et al.* (2010). "It's not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, **4**, 1–13.

Levens, S.M., Phelps, E.A. (2010). Insula and orbital frontal cortex activity underlying emotion interference resolution in working memory. *Journal of Cognitive Neuroscience*, **22**(12), 2790–803.

Li, W., Mai, X., Liu, C. (2014). The default mode network and social understanding of others: what do brain connectivity studies tell us. *Frontiers in Human Neuroscience*, **8**, 74.

Lin, N., Bi, Y., Zhao, Y., Luo, C., Li, X. (2015). The theory-of-mind network in support of action verb comprehension: evidence from an fMRI study. *Brain and Language*, **141**, 1–10.

Lin, N., Wang, X., Xu, Y., *et al.* (2017). Fine subdivisions of the semantic network supporting social and sensory-motor semantic processing. *Cerebral Cortex*, **15**, 1–12.

Mar, R.A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, **62**, 103–34.

Nee, D.E., Jonides, J., Berman, M.G. (2007). Neural mechanisms of proactive interference-resolution. *NeuroImage*, **38**(4), 740–51.

Nelken, I., Ahissar, M. (2006). High-level and low-level processing in the auditory system: the role of primary auditory cortex. In: Divenyi, P., Greenberg, S., Meyer, G., editors. *Dynamics of Speech Production and Perception* (pp. 343–53), Amsterdam: IOS Press.

Nonyane, B. a S., Theobald, C.M. (2007). Design sequences for sensory studies: achieving balance for carry-over and position effects. *British Journal of Mathematical and Statistical Psychology*, **60**(2), 339–49.

Norman-Haignere, S., Kanwisher, N.G., McDermott, J.H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, **88**(6), 1281–96.

Obleser, J., Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, **13**(1), 14–9.

Olson, I.R., McCoy, D., Klobusicky, E., Ross, L.A. (2013). Social cognition and the anterior temporal lobes: a review and theoretical framework. *Social Cognitive and Affective Neuroscience*, **8**(2), 123–33.

Papoušek, M., Bornstein, M.H., Nuzzo, C., Papoušek, H., Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, **13**(4), 539–45.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, **36**(4), 767–76.

Paulmann, S. (2016). The neurocognition of prosody. In: Hickok, G., Small, S. L., editors, *Neurobiology of Language* (pp. 1109–20), Amsterdam: Elsevier.

Pobric, G., Lambon Ralph, M.A., Zahn, R. (2016). Hemispheric specialization within the superior anterior temporal cortex for social and nonsocial concepts. *Journal of Cognitive Neuroscience*, **28**(3), 351–60.

Pourtois, G., Schettino, A., Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: what is magic and what is not. *Biological Psychology*, **92**(3), 492–512.

Prieto, P., Estrella, A., Thorson, J., Vanrell, M.D.M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, **39**(2), 221–57.

Rauschecker, J.P., Scott, S.K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, **12**(6), 718–24.

Ross, L.A., Olson, I.R. (2010). Social cognition and the anterior temporal lobes. *NeuroImage*, **49**(4), 3452–62.

Roy, A.K., Shehzad, Z., Margulies, D.S., *et al.* (2009). Functional connectivity of the human amygdala using resting state fMRI. *NeuroImage*, **45**(2), 614–26.

Said, C.P., Baron, S.G., Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience*, **21**(3), 519–28.

Sammler, D., Grosbras, M.H., Anwander, A., Bestelmeyer, P.E.G., Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, **25**(23), 3079–85.

Sander, D., Grafinan, J., Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences*, **14**, 303–16.

Schirmer, A., Kotz, S. a. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, **10**(1), 24–30.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, **42**, 9–34.

Searle, J.R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press.

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

Skipper, L.M., Ross, L.A., Olson, I.R. (2011). Sensory and semantic category subdivisions within the anterior temporal lobes. *Neuropsychologia*, **49**(12), 3419–29.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., Noveck, I. a. (2012). Neural evidence that utterance-processing entails mentalizing: the case of irony. *NeuroImage*, **63**(1), 25–39.

Sripada, C.S., Angstadt, M., Banks, S., Nathan, P.J., Liberzon, I., Phan, K.L. (2009). Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *NeuroReport*, **20**(11), 984–9.

Stolier, R.M., Freeman, J.B. (2017). A neural mechanism of social categorization. *Journal of Neuroscience*, **37**, 5711–21.

Suzuki, S., Niki, K., Fujisaki, S., Akiyama, E. (2011). Neural basis of conditional cooperation. *Social Cognitive and Affective Neuroscience*, **6**(3), 338–47.

Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H. (2005). Understanding and sharing intention: the origins of cultural cognition. *Behavioral and Brain Sciences*, **28**(5), 675–735.

van Ackeren, M.J., Casasanto, D., Bekkering, H., Hagoort, P., Rueschemeyer, S.-A. (2012). Pragmatics in action: indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience*, **24**(11), 2237–47.

Van Overwalle, F. (2009). Social cognition and the brain: a meta--analysis. *Human Brain Mapping*, **30**(3), 829–58.

Vuilleumier, P., Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia*, **45**(1), 174–94.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., Bara, B.G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, **16**(10), 1854–63.

Walter, H., Ciaramidaro, A., Adenzato, M., *et al.* (2009). Dysfunction of the social brain in schizophrenia is modulated by intention type: an fMRI study. *Social Cognitive and Affective Neuroscience*, **4**(2), 166–76.

Wang, S., Yu, R., Tyszka, J.M., *et al.* (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications*, **8**, 14821.

Warren, J.E., Wise, R.J.S., Warren, J.D. (2005). Sounds do-able: auditory–motor transformations and the posterior temporal plane. *Trends in Neurosciences*, **28**(12), 636–43.

Weiller, C., Bormann, T., Saur, D., Musso, M., Rijntjes, M. (2011). How the ventral pathway got lost—and what its recovery might mean. *Brain and Language*, **118**(1-2), 29–39.

Wiethoff, S., Wildgruber, D., Grodd, W., Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *Neuroreport*, **20**(15), 1356–60.

Wong, C., Gallate, J. (2012). The function of the anterior temporal lobe: a review of the empirical evidence. *Brain Research*, **1449**, 94–116.

Zahn, R., Moll, J., Krueger, F., Huey, E.D., Garrido, G., Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(15), 6430–5.

Zahn, R., Moll, J., Paiva, M., *et al.* (2009). The neural basis of human social values: evidence from functional MRI. *Cerebral Cortex*, **19**(2), 276–83.