Vocomotor and social brain networks work together to express social traits in voices

Stella Guldner[1,2,3*], Frauke Nees[1,4] and Carolyn McGettigan[3,5*]

[1] Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2] Graduate School of Economic and Social Sciences, University of Mannheim, Mannheim, Germany

[3] Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK

[4] Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany

[5] Department of Psychology, Royal Holloway, University of London, Egham, UK

Corresponding authors:

Carolyn McGettigan

Department of Speech, Hearing and Phonetic Sciences

University College London

 2 Wakefield Street

London WC1N 1PF, UK

Telephone: +44 (0)20 7679 4075

Email: c.mcgettigan@ucl.ac.uk

Stella Guldner

Institute of Cognitive and Clinical Neuroscience

Central Institute of Mental Health

J5, 68159 Mannheim, Germany

Telephone: +49 (0) 621 1703 6314

Email: stella.guldner@zi-mannheim.de

1   Abstract

2   Voice modulation is important when navigating social interactions – tone of voice in a

3   business negotiation is very different from that used to comfort an upset child. While

4   voluntary vocal behaviour relies on a cortical vocomotor network, social voice modulation

5   may require additional social cognitive processing. Using functional MRI, we investigated

6   the neural basis for social vocal control and whether it involves an interplay of vocal control

7   and social processing networks. Twenty-four healthy adult participants modulated their voice

8   to express social traits along the dimensions of the social trait space (affiliation and

9   competence), or to express body-size (control for vocal flexibility). Naïve listener ratings

10   showed that vocal modulations were effective in evoking social trait ratings along the two

11   primary dimensions of the social trait space. Whereas basic vocal modulation engaged the

12   vocomotor network, social voice modulation specifically engaged social processing regions

13   including the medial prefrontal cortex, superior temporal sulcus and precuneus. Moreover,

14   these regions showed task-relevant modulations in functional connectivity to the left inferior

15   frontal gyrus, a core vocomotor control network area. These findings highlight the impact of

16   the integration of vocal motor control and social information processing for socially

17   meaningful voice modulation.

18

19   *Keywords*: fMRI, social communication, social traits, vocal control, voice production

20

21

22

23

1   Vocomotor and social brain networks work together to express social traits in voices

2   Similar to the impression management we perform when we dress up for a job interview, can

3   we influence how we are perceived by volitionally modulating our voice? Despite the

4   importance of the voice for social judgement formation and the voice being a fundamentally

5   social behaviour (McGettigan 2015), we understand little about the cognitive and

6   neurobiological underpinnings of expressing social information in voices. Although the voice

7   remains relatively stable within its anatomical boundaries throughout adult life (Pisanski,

8   Fouquet, et al. 2016), vocal behaviour is adaptable to specific social contexts and expressing

9   socially relevant information (reviewed by Pisanski et al., 2016). Consequently, the voice

10   carries a multitude of information about a speaker, such as their emotional state and

11   personality traits (Banse and Scherer 1996; Krauss et al. 2002; Sauter et al. 2010; Pisanski,

12   Cartei, et al. 2016; Oleszkiewicz et al. 2017). Not surprisingly, vocal information is

13   spontaneously used by listeners to infer the intentions of an interlocutor (Hellbernd and

14   Sammler 2016), e.g. to judge how trustworthy or dominant a speaker is (McAleer et al.

15   2014). These judgements can be represented in a social voice space (McAleer et al. 2014),

16   showing spontaneous attribution of traits signalling approach or avoidance, i.e. how likeable

17   or dislikeable a speaker might be, and how socially potent they are, i.e. how intelligent they

18   sound. The social voice space is stable across cultures (Baus et al. 2019) and related to

19   specific acoustic modulation patterns. Pitch contour, for instance, is closely associated with

20   ratings of trustworthiness (Belin et al. 2017; Ponsot et al. 2018), while an interaction of both

21   pitch and intensity measures has been related to expressing hierarchy or confidence in the

22   voice (Ko et al. 2014; Jiang and Pell 2017).

23   The two dimensions of the social voice space, namely affiliation (also termed warmth

24   or trustworthiness) and competence (also termed dominance or confidence) represent the axes

25   of the social trait space (Fiske, Cuddy, & Glick, 2007; Harris & Fiske, 2007), parallel to the

1 previously reported social space of faces (Todorov et al. 2005; Todorov, Said, et al. 2008).

2 Receiving beneficial judgements in the social space is important to achieve successful

3 interactions: previous work suggests that vocal cues might be an important contributor to this,

4 for example predicting positive outcomes in a job interview (Schroeder and Epley 2015) or a

5 political election (Pavela Banai et al. 2017). Dynamic voice changes can be observed

6 spontaneously in response to external cues (e.g. raising the volume of one's voice in a noisy

7 environment), as well as intentionally in response to internal goals (e.g. trying to impress a

8 panel at an interview). Related to the latter context, vocal control describes the capacity to

9 perform goal-directed and voluntary modulation of suprasegmental speech characteristics

10 during voice production. Through vocal control, speakers can influence how old (Skoog

11 Waller & Eriksson, 2016), or how feminine versus masculine they are perceived (Cartei,

12 Cowles, Banerjee, & Reby, 2014). Moreover, immediate social information can also be

13 communicated through controlled vocal modulations, such as social emotions (Morningstar et

14 al. 2017). Such voluntary voice modulation to express social information, i.e. *social vocal*

15 *control*, can directly impact listeners, who make use of vocal information to make

16 spontaneous social trait judgments. Hughes and colleagues (2014) recorded speakers' voice

17 modulations to express traits, such as dominance, and presented them to naïve listeners.

18 Compared to the speakers' normal (i.e. non-modulated) voice recordings, dominant voices,

19 for instance, were indeed rated as higher in dominance than normal voices. This study

20 suggests that social vocal control presents an effective interpersonal tool, which can be

21 instrumental in eliciting beneficial social judgements. However, Hughes and colleagues

22 (2014) only ever obtained perceptual ratings on the *intended* trait for each modulated voice

23 category (e.g. speech intended to sound dominant was only rated for dominance and not, for

24 example, for trustworthiness). Hence, the specificity, and therefore the potency, of social

25 vocal modulations remains unclear.

1   Exerting vocal control has been shown to rely on a fronto-parietal vocomotor control

2   network (VMN) between the IFG pars triangularis/opercularis, supplementary motor area

3   (SMA), the supramarginal gyrus (SMG), insula, superior temporal cortex (STC), anterior

4   cingulate cortex (ACC), basal ganglia (BG) and cerebellum (reviewed by Pisanski, Cartei, et

5   al., 2016; Simonyan & Horwitz, 2011). The IFG has a crucial role in speech motor control

6   during vocalization. It is thought to be a central executive and primary input region for

7   voluntary voice production (Hage and Nieder 2016), representing speech sound maps for

8   feed-forward vocal control, particularly in opercular parts of left IFG (Tourville and Guenther

9   2011). Thus, it provides input to primary motor cortex, which in turn engages a cortical and

10   subcortical network to exert control over vocal production (Simonyan and Horwitz 2011).

11   To our knowledge, no study has specifically targeted the neurobehavioural

12   mechanisms involved in *social* vocal expression. Some insight comes from studies

13   investigating vocal modulation to express affect. The expression of such affective

14   vocalizations has been proposed to rely on the interaction of a dual-pathway system

15   consisting of the neocortical regions of the VMN and a phylogenetically older network of

16   subcortical brain structures such as the basal ganglia and the amygdala (Ackermann et al.

17   2014; Hage and Nieder 2016). In line with this, voluntary affective vocal expression engages

18   both vocomotor areas related to volitional expression as well as areas related to processing

19   affect, such as the IFG, BG, ACC and STC and Amygdala (Barrett et al. 2004; Aziz-Zadeh et

20   al. 2010; Laukka et al. 2011; Pichon and Kell 2013; Frühholz et al. 2015; Klaas et al. 2015;

21   Belyk and Brown 2016; Mitchell et al. 2016; Klasen et al. 2018). This interplay of affect

22   processing streams and the vocomotor network therefore suggests that some informational

23   integration, is necessary to achieve the successful expression of affect in the voice.

24   In line with this, studies investigating the association between social traits and voice

25   production point to an interaction of social processing areas with vocal motor processing

1   areas during socially meaningful voice changes. Klasen and colleagues (2018) showed that

2   activation in the right superior temporal cortex (STC), was modulated by the social context

3   during emotional voice production, suggesting that areas associated with domain general

4   social processing (social brain network; SBN) might be involved in expressing socially

5   relevant information in voices during speech, as it is for perception of social traits in voices

6   (Hellbernd and Sammler 2018). Another study of speech production that directly manipulated

7   third person vocal identity expression (i.e. impersonations) reported activated regions in the

8   right superior temporal sulcus (STS) that were functionally connected to left IFG during

9   voice production (McGettigan et al. 2013). In another study, actors gave improvised, but

10   covert, answers to questions about themselves while in role. In line with the previous studies,

11   responding in the first person of a fictional character (during acting) also engaged right

12   posterior STS regions in addition to a network of ventral (vmPFC) and dorsal medial

13   prefrontal regions (dmPFC) and precuneus, which are all implicated in the SBN (Brown,

14   Cockett, & Yuan, 2019). Although less specifically targeting vocal modulation in their task,

15   the latter study is one of the few that offers some insight into intentional first person trait

16   modulation, and supports the idea that SBN regions might also be instrumental in achieving

17   voluntary expressions of identity in the voice.

18   Social vocal control is an intentional goal-directed behaviour that requires a socially

19   beneficial expression of self-related traits. To be successful, social vocal control should

20   therefore involve some form of trait processing. Together with the medial prefrontal cortex

21   (mPFC), the posterior portion of the STS (or temporo-parietal junction; TPJ) is an important

22   and domain-general contributor to the social brain network (Van Overwalle 2009; Bzdok et

23   al. 2012; Schurz et al. 2014) and is engaged during evaluative judgments of affective

24   information in voices (Dricu and Frühholz 2016). The mPFC is reliably activated during tasks

25   that require mentalizing (Schurz et al. 2014) and has been proposed to subserve social trait

1   judgment along the dimension of the social trait space (Harris et al. 2005; Harris and Fiske

2   2007; Ma et al. 2013, 2016; Van Overwalle et al. 2016), particularly in ventral parts (Harris

3   and Fiske 2007; Van Duynslaeger et al. 2007; Ma et al. 2014, 2016; Tavares et al. 2015; Van

4   Overwalle et al. 2016). Moreover, mPFC is specifically engaged during tasks requiring

5   psychological self-representation, possibly reflecting emotional or evaluative processing of

6   the conceptual self. Compared to other-referential tasks, self-referential tasks are reliably

7   associated with increased activation in SBN areas, such as mPFC, bilateral STG, precuneus,

8   and TPJ (Hu et al. 2016). In fact, mPFC and pSTS/TPJ are involved in trait processing both

9   when social traits are in reference to oneself or to another person (Nicolle et al. 2012). While

10   the mPFC represents rather long-term, static trait information about others (and likely, the

11   self), the pSTS/temporo-parietal junction is involved in rapid, short-term intention and goal

12   attribution (Saxe and Powell 2006; Ma et al. 2012) both for verbal and nonverbal information

13   (Redcay 2008; Shultz et al. 2012; Redcay et al. 2016). In summary, social vocal control to

14   express beneficial traits might entail an integration of vocomotor and social trait processing

15   areas in relation to the self. Although social trait perception in voices has recently gained

16   attention (e.g. Hellbernd & Sammler, 2018), no study to our knowledge has tested this

17   directly in voice production.

18   Lastly, although some evidence points to the representation of a common trait code in

19   the ventral mPFC (Van Overwalle et al. 2016), the question of whether the two main

20   dimensions of the social trait space engage separable neural regions remains controversial.

21   While the affiliative traits (e.g. warmth, trustworthiness) are associated with processing in

22   ventral mPFC, competence evaluations have been shown to additionally engage the

23   precuneus (Ma et al. 2016). In the present study, both competence and affiliative traits were

24   expressed volitionally in the voice in our social vocal control task, allowing insights on the

25   neural representation of these two social space dimensions.

1   In summary, although speaking can be understood as a goal-directed social behaviour,

2   little is known about how social traits are encoded in the voice. The current study addressed

3   this gap in the literature, posing two research questions on the issue. First, we tested whether

4   social vocal control would be effective in evoking percepts that vary along the dimension of

5   the social trait space: affiliation and competence. Here, we tested in particular beneficial

6   social traits, i.e. sounding likeable or hostile, and sounding intelligent to navigate

7   interactions. This is in alignment with the spontaneous perception of voices on the social

8   voice space (McAleer et al. 2014) and comparable to commonly expressed social information

9   in everyday social interactions. Second, we aimed to illuminate the functional neural

10   correlates supporting social vocal control to navigate this social space, i.e. to express social

11   traits. Based on the literature on neural correlates of basic vocal control, we predicted

12   significant changes in neural activation during vocal modulation in the VMN (including the

13   pars triangularis and pars opercularis of the left IFG as a central vocal control area, as well as

14   supplementary motor area, basal ganglia, cerebellum, and insula). For social vocal control,

15   we expected a higher engagement of areas involved in modality-independent social trait and

16   self-referential trait processing, such as pSTS/TPJ and mPFC. To achieve social vocal

17   control, we expected that the core vocal control network (left IFG) would be functionally

18   connected to social processing areas during online social voice modulation. According to

19   accounts of the social trait space, we further explored whether social vocal control to express

20   traits along the two dimensions of affiliation and competence would rely on differential

21   engagement of social processing areas and lastly, whether there would be differential

22   functional connectivity profiles with left IFG for these two dimensions of the social trait

23   space.

1   **Methods**

2   **Participants.** Twenty-four right-handed, native British English speakers ($M_{age} = 21.04$

3   SD=3.26, 3 male) participated in this experiment. All participants had normal or corrected-to-

4   normal vision and reported no history of hearing, language, neurological or psychiatric

5   disorders. Volunteers were recruited from the participant pool at the Department of

6   Psychology at Royal Holloway, University of London and received 30 GBP as

7   reimbursement. All participants provided their full informed and written consent prior to

8   participation according to the Declaration of Helsinki (1991; p. 1194). This study was approved

9   by the research ethics committee of Royal Holloway, University of London (587-2017-10-24-

10   14-50-UXJT010).

11   **Social vocal control task.** The main experimental task consisted of a social vocal control

12   task in which participants were asked to express social and non-social traits in the voice. The

13   social traits were selected to represent the two principal dimensions of the social trait space

14   (Fiske et al. 2007), namely affiliation (also warmth/trustworthiness) and competence (also

15   referred to as dominance; see Belin, Boehme, & McAleer, 2017; McAleer, Todorov, & Belin,

16   2014 for prior work on trait dimensions in voices). Social traits included vocally expressed

17   intelligence (competence), and likeability and hostility (affiliation dimension), thus spanning

18   the social voice space reflecting spontaneous social appraisal of neutral voices (McAleer et

19   al. 2014). Modulating the voice to express a large body size, as well as speaking in non-

20   modulated "normal voice", were implemented as control conditions (see supplementary

21   materials S1). A body-size modulation was chosen as a control condition because it demands

22   substantial vocal tract manipulations (e.g. larynx lowering, changes in vocal fold tension),

23   without having direct social trait implications (Pisanski, Mora, et al., 2016). Exemplars

24   consisted of four two-syllable, five-letter pseudowords with a C-V-C-V-C (C=consonant, V=

1   vowel) phonotactic structure (i.e. *belam*, *lagod*, *minad*, and *namil*; Frühholz et al. 2015).

2   These exemplars have previously been used in tasks involving voice production with

3   affective content (Frühholz et al., 2015).

4   **Design and procedure**. Prior to the main experiment, participants completed a short training

5   task, during which they were introduced to the exemplars and social traits to be expressed and

6   could familiarize themselves with the social vocal control task. In the scanner, participants

7   then completed the task, which consisted of 4 runs of 150 trials, including 30 rest trials. Each

8   run consisted of a randomized order of the 5 vocal modulation conditions (normal / large /

9   hostile / likeable / intelligent) paired with one of the 4 exemplars. Each exemplar and

10   condition combination was repeated 6 times over the course of a run. Out of these, 3

11   repetitions were Go trials and 3 were No-Go trials. Go and No-Go trials were presented in

12   randomized order, with the restriction of a maximum of three No-Go trials in a sequence.

13   Within each trial, participants were presented with a social target trait and a fixation cross for

14   two seconds. During this time they were asked to prepare to express the target trait in their

15   voice. In Go trials, the exemplar was then shown at the position of the fixation cross for 1.5

16   seconds. At the beginning of this silent gap the exemplar appeared on the screen and

17   participants were asked to vocalize the exemplar while expressing the target trait. During No-

18   Go trials the fixation cross remained on the screen for the duration of the silent gap and no

19   exemplar was presented (see Figure 1). We included both Go and No-Go trials in the task, to

20   filter neural activation specifically related to active, ongoing voice production (Go) and

21   thereby exclude any other task-related cognitive processing effects (present in Go and No-

22   Go). Importantly, participants prepared to speak in both trials, but only received a target word

23   in Go trials.

24   Vocal recordings of speech in Go trials were recorded using an in-scanner MR-

25   compatible microphone (Opto-acoustics, FOMRI-III). Visual cues were presented using the

1 Psychophysics toolbox (Brainard, 1997, Kleiner, Brainard & Pelli, 2007) in Matlab (2014a,

2 the Mathworks, Natick, MA) – these were projected onto a screen at the back of the scanner

3 bore and viewed via a mirror on the headcoil. The total scanning time was around 50 minutes.

4   After scanning, participants were re-invited to rate their own performance on the

5 social vocal control task, using the voice recordings acquired during the scanning session. To

6 do this, they were presented with their voice recordings blocked by the expressed target trait.

7 Recordings were presented in a soundproof booth via Sennheiser Headphones HD306, using

8 the Psychophysics toolbox (Brainard 1997, Kleiner, Brainard & Pelli, 2007) in Matlab

9 (2014a, the Mathwork, Natick, MA). For each recording the participant was re-invited

10 approximately 1 week later and asked to evaluate how strongly they had expressed the social

11 trait on a 7-point Likert scale ranging from not at all (=1) to very much (=7). Based on these

12 subjective ratings, we selected one recording for each trait category, for each participant that

13 had received the participant's own maximal rating for that trait. This was done to ensure that

14 independent evaluations of performance (i.e. with naïve listeners) would be based on sounds

15 where the original participants had felt confident in their accurate expression of the target

16 trait. In cases where multiple recordings had the same maximal rating, we selected one token

17 at random.

18

19   *Figure 1 about here*

20

21 **Naïve ratings.** For each participant the highest rated recording for each trait category was

22 then presented to naïve listeners to obtain a performance index. For this task, the normal

23 voice recordings were intensity normalized across all 24 speakers. The resulting

24 normalization parameter was then used to normalize trait recordings within each speaker: this

25 preserved within-speaker differences between the traits while maintaining average intensity

1 levels between speakers. The normalized recordings were then presented to 24 naïve listeners

2 (M_{age}=19.92, SD=1.47, 4 male). Listeners were recruited through at the Department of

3 Psychology at Royal Holloway, University of London. All raters signed their informed

4 consent prior to participation and received 5 GBP as reimbursement for their participation.

5   To reduce experiment duration and avoid fatigue effects, each rater heard only a

6 subset of 10 speakers. The subset of speakers was counterbalanced to ensure that each

7 speaker was heard by at least 10 different listeners. Recordings were presented in randomized

8 order within 7 blocks. Each block consisted of one of the 4 relevant social trait modulations

9 to be rated (hostility, likeability, intelligence and body size), or an additional stimulus

10 property (i.e. arousal, valence, authenticity). One recording of each speakers' normal voice

11 was included in the task, since we were interested how the trait ratings induced by the vocal

12 change differed from the normal voice of each speaker. Using this approach, each listener

13 rated each recording on all traits over the course of the blocks: Ratings were given on 7-point

14 Likert scales measuring how strongly the heard voices expressed a given trait, ranging from

15 not at all (=1) to very (=7). For the valence ratings, the scale ranged from very negative (=-3)

16 to very positive (=3). Arousal ratings were rated on a Likert-scale ranging from very sleepy

17 (=1) to very alert (=7). Listeners also rated each recording on perceived authenticity –

18 however, due to some uncertainty about how participants interpreted this scale we have

19 chosen not to include the authenticity ratings in the analyses reported here. The order of

20 blocks was randomized between listeners. Stimuli were presented using the Psychophysics

21 Toolbox (Brainard 1997, Kleiner, Brainard & Pelli, 2007) in Matlab (2014a, the Mathwork,

22 Natick, MA) on Sennheiser Headphones (Sennheiser U.K. Ltd, Marlow, UK) in a soundproof

23 booth.

24

25

1   **fMRI acquisition and analysis**

2   **Task-based fMRI.** Functional brain images were acquired on a 3T Siemens TIM Trio

3   scanner with a 32 channel headcoil, using a rapid–sparse event-related 3D echo-planar

4   imaging (EPI) sequence (32 axial slices, slice gap 25%, resolution 3x3x3mm^2, flip angle 78°,

5   matrix 64 x 64, TE: 30 msec, TR: 3.5 sec, TA: 2 sec). A 3D T1-weighted MP-RAGE scan

6   was acquired for EPI image alignment and spatial normalisation (voxel size 1 mm isotropic;

7   flip angle 11°; TE 3.03 ms; TR 1830 ms; image matrix 256 x 256). Analysis was conducted

8   in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). Preprocessing steps included spatial

9   realignment, segmentation, co-registration, normalization (functional images were resampled

10   to a voxel size of 2x2x2mm) and smoothing (FWHM=8mm). 1st Level general linear models

11   included the conditions as regressors and subjective ratings as parametric modulators, which

12   were analysed in the framework of one-sample T-tests at the second level. We used a

13   significance threshold of $p<.001$ for second level tests, uncorrected for multiple comparisons.

14   To ensure a type 1 error of $p=.001$ at the individual voxel level and a threshold of $p=.05$

15   corrected for multiple comparisons at the cluster level, a cluster extent threshold was

16   computed for each contrast. This threshold was determined using 1,000 Monte-Carlo

17   simulations based on whole-brain fMRI activation, as described elsewhere (Slotnick et al.

18   2003). The resulting clusters were labelled based on the location of each peak activation

19   using the in-built Neuromorphometrics and the automated anatomical labeling (AAL) atlas in

20   SPM12. For illustrations, parameter estimates were extracted from significant clusters in the

21   group maps using the MarsBar Toolbox in SPM12 (Brett et al. 2002).

22   **Psychophysiological Interactions**. Lastly, we investigated the integration of social

23   processing with vocal motor processing reflected in task-related connectivity changes during

24   exerting social vocal control with the core vocal motor area in left IFG. To this end, we

13

1   computed psychophysiological interactions (PPI) in SPM12. To isolate the individual core

2   vocal motor area, we constructed VOIs (volumes of interest) based on the peak activation of

3   the 2nd level group contrast of Go > No-Go trials in left IFG. A sphere of 10 mm radius was

4   constructed around the peak of this activation map ([-52 12 24]) and used as a mask image to

5   search for peak activation on a single-subject level with a voxel height threshold of $p < .001$.

6   We then defined VOIs on the peak activation for each participant and built individual spheres

7   with radius of 6 mm (i.e. 3 voxels) around this peak coordinate. The first eigenvariate of the

8   functional MRI signal change was then extracted from the VOIs and the mean time course

9   was multiplied by the task regressors. In separate PPI analyses, the task regressors were based

10   on the following contrasts: 1) Go > No-Go for social modulation trials, and 2) affiliation

11   (likeable and hostile) > competence (intelligence) trials. We added these interaction terms for

12   each model as regressors to the 1st level models, along with the deconvolved source signal of

13   the VOI and the task regressors. Contrasts between Go and No-Go trials for the relevant

14   conditions (e.g. social vocal control) were chosen, because we were interested in exploring

15   mechanisms specific to voice production, while subtracting out other general task-related

16   effects that were common to both task types (e.g. imagining a social scene). Thus, the

17   contrast of Go > No-Go allowed us to target the mechanisms specifically involved in

18   achieving ongoing social voice production.

19   **Statistical analysis of rating data**

20   **Vocal modulation performance.** The ratings obtained from naïve listeners were analysed in

21   R (http://www.R-project.org/). To assess the success of the voice manipulations to express

22   social traits, we calculated the average change in naïve ratings for modulated voices relative

23   to the normal voice samples, for the intended trait (e.g. comparing "intelligent" ratings for the

24   normal and "intelligent" trials) and for the other traits (e.g. comparing "intelligent" ratings for

1 the normal and the "likeable" trials) individually for each speaker and each trait. Thus, we

2 obtained the change in mean ratings, henceforth Δ - ratings, from each speaker's normal

3 voice as a function of each expressed trait (e.g. intelligent ratings of "intelligent" modulation

4 – intelligence ratings for the normal voice), and for each trait condition, allowing us to

5 measure the effectiveness of volitional social trait expression, i.e. the sensitivity and

6 specificity with which voice changes evoked the intended social trait percepts in listeners.

7 The Δ - ratings were then analyzed in the framework of linear mixed effects models for each

8 trait separately ('lme4' package; Bates, Mächler, Bolker, & Walker, 2015). Each model

9 included the expressed trait as a fixed effect term and speaker as a random intercept to

10 account for within subject variation. Likelihood ratio tests were performed to test the effect of

11 trait expression on the Δ - ratings, by comparing the models with fixed effects to the null

12 models with only the random intercepts. We implemented planned treatment contrasts with

13 the congruent trait rating as a reference, to test directly whether sounds from the congruent

14 trait condition received significantly higher trait ratings than all other voice modulation

15 conditions. Statistical significance for all models was set at a Bonferroni corrected

16 significance level of $p = 0.013$ (for 4 comparisons within each model).

17 We also contrasted ratings on arousal and valence across the social trait modulations

18 in the framework of linear mixed effects models, using the expressed trait as a fixed effect

19 term and the speaker as a random intercept. Here, the mean ratings were entered into the

20 model, because we aimed to statistically contrast differences in comparison to the normal

21 voice. Therefore, we again implemented planned treatment contrasts in these models with the

22 ratings obtained for the normal voice as reference. Again, statistical significance for all

23 models was set at a Bonferroni corrected significance level of $p = 0.013$ (for 4 comparisons

24 within each model). Statistical assumptions for all implemented linear mixed effects models

25 were tested and met.

1   **Social trait space dimensions in voice modulation.** To test the differentiation of the trait

2   categories based on the multivariate naïve ratings (Δ - ratings, which measures the change in

3   ratings relative to the normal voice on the different traits, as well as arousal and valence), we

4   computed a linear discriminant analysis in R ('MASS' package; Venables & Ripley, 2002).

5   The final model included the trait category (i.e. likeable, hostile, intelligent and large) as

6   dependent variable and the mean – centered change in naïve ratings for the modulated voices

7   from the normal voice (Δ - ratings) on all trait scales, as well as on arousal and valence, as the

8   predictor variables. This allowed us to explore the contribution of individual trait and control

9   (i.e. arousal and valence) ratings to discriminant functions and test whether these would be

10   differentiable along the conceptual distribution of the affiliation and competence dimension

11   of the social trait space (Fiske et al. 2007; McAleer et al. 2014; Belin et al. 2017). In other

12   words, we tested whether changes in ratings evoked by modulating the voice represent

13   exaggerations within the trait space.

14   <div align="center">**Results**</div>

15   **Behavioural results**

16   **Subjective ratings.** Subjective performance (as assessed by the talkers themselves) differed

17   among the social traits expressed in the social vocal control task, $\chi^2(4)=40.26$, $p<001$.

18   Planned contrasts showed that likeable voice modulations (M=5.37, SD=.64) were perceived

19   as equally successful as the normal (M=5.68, SD=0.72) voice expression, t=-1.89, p=.06.

20   That is, normal voice trials sounded as normal as likeable modulations sounded likeable. All

21   other trait modulations were perceived as less intensely expressed in the voice, yet above the

22   midpoint of the 7-point Likert scale (hostile: M=5.20, SD=0.61, intelligent: M=4.59,

23   SD=0.60, large: M=4.97, SD=0.76, all ts < -2.913, all ps<.01). This indicates that participants

24   felt they were able to do the task, in spite of differences in performance between the traits.

1   The subjective speaker rating for each recording was included as a parametric modulator in

2   the fMRI analysis for the respective condition regressor to account for differences in

3   functional activation related to task difficulty.

4   **Vocal modulation performance.** For analysis of the naïve listener ratings, all voices (normal

5   and modulated) were rated on all trait scales (likeable, hostile, intelligent, large). There was a

6   significant relative change in the naïve listener ratings evoked from social vocal modulation

7   in all trait categories relative to normal voice recordings (Δ - ratings), indicating that social

8   vocal modulation was perceivable by the listeners (all χ^2s(3)>29.40, ps<001, see Figure 2).

9   Planned contrasts showed that modulated voices expressing likeability were rated as

10 sounding significantly more likeable than voices expressing other traits (all bs<-1.18, all

11 ts(72)<-4.87, all ps<.001). Likeable voices were also perceived as more positive (b=-1.11,

12 t(96)=5.29, p<.001) and higher in arousal than normal voices (b=0.90, t(96)=3.53, p<.001).

13   Hostile voice modulation lead to significantly greater relative ratings of hostility, than

14 did likeable modulations (b=-2.68, t(72)=-10.11, p<.001), or intelligent voices (b=-1.76,

15 t(72)=-6.71, p<.001). Voices expressing larger body-sizes were perceived as similarly hostile

16 as hostile voices (b=-0.15, t(72)=-0.55, p=.58). Hostile voices were also rated as more

17 negative (b=-1.39, t(96)=-6.60, p<.001) and higher in arousal than normal voices (b=1.35,

18 t(96)=5.01, p<.001).

19   Intelligent voices were perceived as sounding more intelligent than hostile voices

20 (b=-1.11, t(72)=-5.39, p<.001) or large voices (b=-0.93, t(72)=-4.51, p<.001). Likeable

21 voices also gave relatively increased intelligence ratings, although marginally less so than

22 intelligent voices (b=-0.40, t(72)=-1.92, p=.059). Intelligent voices were perceived as higher

23 in arousal (b=1.06, t(96)=3.93, p<.001) but similarly neutral in valence as normal voices (all

24 bs<0.22, all ts(96)<-1.13, all ps>.05).

1   Lastly, voices modulated to sound larger induced a positive change in size ratings that was

2   significantly higher than for likeable voices (b=-1.80, $t(72)$=-8.60, p<.001) or intelligent

3   voices (b=-1.17, $t(72)$=-5.59, p<.001). Large voices were rated similarly large as hostile

4   voices (b=-0.18, $t(72)$=-0.88, p=.383), but were perceived as higher in arousal than hostile

5   voices (b=2.04, $t(96)$=7.57, p<.001). Large voice modulations were also perceived as more

6   negative than normal voices (b=-1.27, $t(96)$=-6.03, p<.001). Taken together, congruent trait

7   ratings (i.e. when the expressed trait and the rated trait coincided) were generally rated

8   significantly higher than incongruent trait ratings (i.e. when expressed traits did not coincide

9   with trait ratings; see Figure 2).

10   *Figure 2 about here*

11

12   **Social trait space dimensions in voice modulation.** Linear discriminant analysis further

13   showed a successful differentiation between social voice modulations based on naïve ratings

14   of traits and ratings of arousal and valence (Wilk's lambda=.355, $F(12,227.8)$=7.06, p<.001).

15   Four recordings were removed as multivariate outliers from the analysis, leaving 92

16   modulation recordings in the model. The model showed an overall classification accuracy of

17   57% (95% CI = 0.46, 0.67), which was significantly above the No Information Rate (NIR =

18   26%, p<.001). Based on the ratings, likeable voices were best differentiated, with high

19   sensitivity (87%) and specificity (83%), whereas large voices (from the body size condition)

20   were least classifiable with a sensitivity of 32% but specificity of 87%. A combination of

21   three linear discriminant functions allowed this classification, whereby the first two linear

22   discriminants explained 96% of total variance. Based on the modest discriminant power of

23   the third function, it was not analyzed further. The first function differentiated best between

24   likeable and hostile voice modulation, accounting for 88% of explained between-group

25   variance. Differentiation based on this discriminant function relied on hostile (b=.33), body-

1 size (*b*=.50) and likeable (*b*=-.19) voice ratings, but not intelligence ratings (*b*=.02). Among

2 the control ratings, valence ratings (*b*=-.45) and arousal (*b*=-.27) also contributed to the

3 differentiation. The second discriminant differentiated between intelligent voices and voices

4 expressing affiliation traits (likeable and hostile voices), accounting for 8% of the between-

5 group variance (see Figure 3). Intelligence (*b*=1.14) ratings loaded on this discriminant in the

6 opposite direction of hostile (*b*=-.32), likeable (*b*=-1.03), as well as body-size ratings (*b*=-

7 .59). Arousal (*b*=.18) ratings also contributed to this function, whereas valence did not

8 (b=.02). Thus, the changes in rating behaviour induced by the vocal modulation can be

9 differentiated relative to the expressed trait and together, reflect the two dimensions of the

10 social trait space affiliation and competence as reported previously (e.g. Fiske et al. 2007;

11 McAleer et al. 2014; see Figure 3).

12

13   *Figure 3 about here*

14

15 **fMRI results**

16 **Group contrasts**. Voice Modulation as opposed to normal voice (all modulation conditions >

17 normal voice (Go trials)) induced changes in functional activation in 5 clusters, showing peak

18 activations in the bilateral insulae, right superior temporal gyrus, left inferior frontal gyrus

19 (IFG: triangular and opercular portions), supplementary motor area (SMA) and the anterior

20 cingulate cortex (ACC), left supramarginal gyrus (SMG) and posterior parts of corpus

21 callosum at a threshold of *p*<.001, minimal cluster size of *k*=61 voxels. In contrast, speaking

22 in a normal voice elicited changes in activation in bilateral inferior parietal cortex (IPC),

23 bilateral middle frontal gyurs (MFG), bilateral middle temporal gyrus (MTG), posterior

24 cingulate cortex (PCC), and left cerebellum (see Figure 4 and Table 1).

25

1   *Figure 4 about here*

2   *Table 1 about here*

3

4

5   Social voice modulations compared to nonsocial voice modulations (likeable ∩

6   hostile ∩ intelligent) > body-size, thresholded at uncorrected *p*<.001, minimum cluster size of

7   *k*=61) induced changes in functional activation in 9 clusters with peak activation in left

8   hippocampus (HC), dorsal and ventral portions of the medial prefrontal cortex (mPFC), left

9   cuneus and precuneus, bilateral lingual gyri, bilateral superior temporal sulci (STS) and

10   bilateral retrosplenial cortex. In contrast, nonsocial voice modulations (body-size > likeable

11   ∩ hostile ∩ intelligent) compared to social voice changes engaged left-lateralized regions in

12   the triangular portions of the IFG (see Figure 5 and Table 2).

13

14   *Figure 5 about here*

15   *Table 2 about here*

16

17

18   During vocal modulation along the affiliation dimension of the social trait space

19   (hostile ∩ likeable > intelligent, thresholded at *p*<.001, minimum cluster size of *k*=58), we

20   found functional activation changes in the left amygdala, the right posterior STS/temporo-

21   parietal junction (TPJ), right SMG, right precentral gyrus and an activation cluster spanning

22   from posterior cingulate cortex to the precuneus. The competence dimension (intelligent >

23   hostile ∩ likeable) induced activation changes in 4 clusters, including a cluster in the left

24   IFG, spanning opercular and triangular portions and into medial frontal gyrus, inferior frontal

25   gyrus pars orbitalis, bilateral superior frontal gyrus, including SMA, as well as a cluster

26   spanning over lingual gyrus bilaterally and left cuneus/calcarine gyrus (see Figure 6 and

27   Table 3).

1   *Figure 6 about here*

2   *Table 3 about here*

3   **PPI.** The PPI analysis revealed changes in functional connectivity of the left IFG with

4   regions in dorsal mPFC, right putamen, left posterior cingulate cortex (PCC) and precuneus,

5   middle cingulate cortex, right posterior insula, right IFG (opercula and triangular portions)

6   and cerebellum during social voice modulation trials (Go > No-Go social voice modulation).

7   Inversely, during No-Go social modulation trials, functional connectivity increased between

8   IFG and left anterior insula (see Figure 7 and Table 4). During trials requiring vocal

9   modulation along the affiliation dimension, we observed changes in functional connectivity

10   between the IFG and clusters in the dorsal striatum (spanning to anterior insula), middle and

11   posterior portions of right STS, and right triangular and orbital portions of the IFG (see

12   Figure 7 and Table 4). There was no meaningful change in functional connectivity between

13   the left IFG and other brain regions during competence voice modulation (see also

14   supplementary materials S3). All contrasts were thresholded at uncorrected $p<.001$, minimal

15   cluster size of 58 voxels.

16

17   *Figure 7 about here*

18   *Table 4 about here*

19

20   **Discussion**

21   The voice is both a dynamic social behaviour and a rich source of information about a person.

22   Successful modulation of the voice to express socially relevant information is an important

23   contributor to achieving interactional and communicative goals. In this study, we found

24   support for our hypotheses: First, we showed that social trait judgements are modulated along

1   the social trait space dimensions as a function of social trait expression in voices. Second,

2   neural activation data showed activation of processing networks related to social trait

3   processing (STS, pSTS/TPJ, mPFC, Precuneus) and vocomotor control (left IFG, SMA,

4   SMG, ACC) during the performance of voluntary, socially-relevant vocal modulation.

5   Finally, functional connectivity analyses suggest an interaction between the left IFG and the

6   social brain network during the performance of social vocal modulations.

7   **Vocal control of social traits expression**

8   Our behavioural data showed that volitional expression of social traits in the voice was not

9   only recognizable as a voice change, but effective in its intention: it led to specific changes in

10   perceptual ratings relevant to the targeted social trait. One previous study has investigated

11   vocal modulation to express trait information (Hughes et al. 2014). In this study, the

12   modulated speech exemplars were sequences of numbers, which were rated by naïve listeners

13   and compared to the rating of the normal voice for the congruent social trait only. We extend

14   these findings, showing evidence for specificity and effectiveness of vocal modulations.

15   Further, we show that social voice modulations during speaking of pseudowords are

16   differentiable for naïve listeners on a multivariate level, and that the change in rating

17   behaviour relative to hearing normal voices was best discriminated on two discriminants

18   relating to the affiliation and competence dimensions of the social trait space. Our data

19   therefore suggest that voluntary social vocal modulations evoke a change in perception of the

20   speaker's voice, which can amplify a social trait rating relative to their normal voice.

21   We observed that vocal modulations to increase perceived body size and hostility

22   were most often perceptually confused compared to the other modulation conditions. This

23   might be driven by the fact that acoustic information related to body size, the body size

24   projection, is an important contributor to perceptions of emotions in voices

25   (Chuenwattanapranithi et al. 2009). Anger in particular, is related to an increased body size

1   projection through vocal tract elongation. Although in this work, we did not ask speakers to

2   express any emotion, we acknowledge that expressing hostility and anger might have similar

3   social implication, i.e. to keep a safe distance from a speaker. Nevertheless speakers reported

4   very differentiable scenarios during the hostile compared to the large voice modulations (e.g.

5   "speaking louder" when trying to increase body-size versus "speaking to somebody I don't

6   like" during expression of vocal hostility). This differentiation was observable in speakers on

7   the neural level, where only the supplementary motor area (SMA) showed overlapping in

8   functional engagement during the two conditions (see supplement S4). This implies

9   differentiable underlying social cognitive processes when modulating the voice to express a

10   large body size versus hostility, in the face of perceptual confusion of body-size and hostility

11   on the receiver end.

12   Trait judgements in the social space have been replicated extensively (reviewed by

13   Fiske et al., 2007) and might reflect domain-general processing of information about the

14   intentions of others, including judgements of personality from voices (McAleer et al. 2014;

15   Baus et al. 2019), faces (Todorov, Said, et al. 2008) or stereotyping of social groups (Fiske et

16   al. 2002). Although the voice has predominately been studied as a vehicle for spoken

17   language, theorists have described speech as actions primarily carrying intentional force

18   (Austin 1975; Grice 1975). In fact, the voice is primarily a tool for interpersonal

19   communication – vocal behaviour mainly occurs when there is an intention to communicate

20   information to others. Thus, the voice is an important source of information about a speaker's

21   intentions. Our data underline the potential social benefit that could arise from successful

22   vocal modulations in interactions with others, whether to express liking, maintain distance, or

23   convey competence.

24   Proficiency in the volitional control of social expression in the voice might be an

25   important contributor to successfully managing impressions in a variety of social situations,

1   from job interviews (Schroeder and Epley 2015) to political campaigns (Pavela Banai et al.

2   2017). As such, it could be a strategic tool with which social opportunists might manage

3   another's impressions to achieve beneficial outcomes. One previous study has explored the

4   effect of affective subcomponents of the psychopathic personality on prosody perception

5   networks (Aziz-Zadeh et al. 2010). This study suggested a positive association between

6   affective empathy levels and functional activation in common motor and perception regions

7   during listening to affective prosody. However, whilst showing decreased sensitivity to socio-

8   affective cues (Blair et al. 2007), individuals scoring high on psychopathy are particularly

9   effective in volitional affective facial expressions (Book et al. 2015) and unimpaired in

10   social-cognitive tasks (Theory of Mind abilities; Winter, Spengler, Bermpohl, Singer, &

11   Kanske, 2017). Whether this extends to the voice - that is, whether trait cognitive versus

12   affective empathy also has an effect on social voice modulation efficiency - remains unclear.

13   Thus, the potential dissociation between perceptual sensitivity to spontaneous social cues as

14   compared to the production of volitional social cues might be an important avenue for future

15   studies.

16   **The VMN in voluntary voice modulation**

17   We report a modulation network engaged in voluntary vocal control in left IFG, ACC, SMA,

18   SMG, STG and insula. This network overlaps largely with the previously reported vocomotor

19   network (VMN) involved in voluntary vocal control (left IFG, SMA, SMG, ACC, STC and

20   insula; Barrett et al., 2004; Golestani & Pallier, 2007; McGettigan et al., 2013; Peschke,

21   Ziegler, Eisenberger, & Baumgaertner, 2012; Pisanski, Cartei, et al., 2016; Reiterer, Hu,

22   Sumathi, & Singh, 2013; Simmonds, Leech, Iverson, & Wise, 2014). Our experimental set-up

23   allowed us to compare specifically which parts of the VMN are centrally engaged in non-

24   social aspects of vocal control, by asking participants to modulate the expression of body

25   size. Here, particularly left IFG (pars triangularis), showed significant changes in activation

1 when speakers produced vocal modulation without social content. The left IFG is a central

2 structure of the volitional vocal control network involved in all aspects of vocalization

3 (laryngeal, orofacial and respiratory control; reviewed by Simonyan & Horwitz, 2011). It is

4 anatomically connected to ACC and the laryngeal and facial motor areas in the primary motor

5 and ventral premotor cortex. Representing sound maps for feed-forward processing

6 (Tourville and Guenther 2011), as well as for motor coordination, l-IFG exerts top-down

7 inhibitory control of the subcortical affective vocal network via the ACC (Hage and Nieder

8 2016). We corroborate the central role of left IFG in voluntary vocal control. Moreover, vocal

9 body-size modulations are typically associated with changes of both fundamental frequency

10 and formant frequencies (Pisanski, Mora, et al. 2016). The triangular part of the l-IFG shows

11 enhanced activation during modulation of prosodic information of the voice (Agnew et al.

12 2017) and is involved in semantic processing of vocal information during speech perception

13 (Gough et al. 2005). Keeping with these findings, our findings support the notion that l-IFG

14 (in particular pars triangularis) provides input into the formation of sound maps for speech

15 that encompass supra-segmental and prosodic vocal meaning.

16 **Neural mechanisms underlying social voice modulation**

17 Our social vocal control task required participants to modulate their voice to express social

18 traits. We hypothesised that this task engages the SBN, a set of regions related to social

19 cognitive functions, including social trait and self-referential processing. In line with our

20 hypothesis, the social vocal modulation conditions (hostile, likeable, intelligent) engaged the

21 dorsal and ventral portions of the medial prefrontal cortex (mPFC), the bilateral superior

22 temporal sulci (STS), left hippocampal formation and precuneus more strongly than the non-

23 social vocal modulation condition (body size). These areas comprise the SBN (Van

24 Overwalle 2009; Schurz et al. 2014), and have been partly implicated in previous studies

25 requiring socially meaningful voice production, during impersonations (McGettigan et al.

1   2013; Brown et al. 2019) or while volitionally modulating the voice within a social contexts

2   (Klasen et al. 2018). In the current study, we show the first evidence for engagement of social

3   processing areas during voluntary voice change to express beneficial social traits.

4   MPFC is a core SBN area involved in domain-general social cognitive processing

5   (Bzdok et al., 2012; Molenberghs, Johnson, Henry, & Mattingley, 2016; Schurz et al., 2014;

6   for review see Van Overwalle, 2009) and mental state inference from vocal cues (Dricu and

7   Frühholz 2016). Importantly, mPFC is suggested to represent social knowledge (Krueger et

8   al. 2009), related to the social trait space (Van Overwalle 2009; Ma et al. 2011, 2012, 2014,

9   2016; Van Overwalle et al. 2016) in reference to others *and* the self (Nicolle et al. 2012; see

10   also meta-analytic evidence of mPFC representation of the psychological self in Hu et al.

11   2016). Our findings corroborate the involvement of mPFC in processing related to social trait

12   knowledge and importantly, extend previous work in showing its involvement in social

13   evaluation, related to both others' and *own actions towards others*. In the present study, we

14   manipulated the modulation of the voice to express social traits, which can be understood as a

15   goal-directed social behaviour (Wolpert et al. 2003). Such voluntary social behaviour has

16   been suggested to rely on mPFC engagement (Thornton et al. 2019), with specific roles for

17   ventral and dorsal regions (Krueger et al. 2009). While dorsal mPFC is thought to represent

18   social goal orientation and execution, ventral mPFC activation reflects self-relevant outcome

19   expectations by modelling social behaviour (Nicolle et al. 2012) based on social trait

20   knowledge (Ma et al. 2014). In support, we found both ventral and dorsal regions in mPFC

21   are engaged during social voice change. Moreover, specifically dorsal mPFC was

22   functionally connected to l-IFG during socially modulated speech, suggesting it may serve as

23   an input region to l-IFG. We therefore suggest that engagement of mPFC in this study might

24   reflect the speaker accessing social trait knowledge to inform the formation of specific trait

25   related vocomotor maps to achieve goal-directed vocal adjustments.

1   Together with mPFC, middle parts of the STS showed enhanced activation during

2   social vocal control conditions. The STC contains both emotion-sensitive (Kreifelts et al.

3   2009) and voice-sensitive (Belin et al. 2000) areas, and similarly to the fusiform gyrus for

4   face processing, has been reported to be involved in voice identity processing (Belin and

5   Zatorre 2003; Schall et al. 2014) and expression (McGettigan et al. 2013). While STG is

6   involved in auditory prediction modelling during vocal control (e.g. Frühholz et al., 2015;

7   conceptualized in the DIVA-Model of speech production by Tourville & Guenther, 2011),

8   STS is reliably involved in inferring communicative intent from observed actions (e.g.

9   Redcay et al., 2016; Schurz et al., 2014). Our current findings suggest that activity in the STS

10   might not only be related to processing intention in others, but communicating one's own

11   social intentions to others, too.

12   Lastly, precuneus involvement is seen in tasks requiring social trait inference (Tavares

13   et al. 2015; Ma et al. 2016; Van Overwalle et al. 2016) and self-referential processing (Hu et

14   al. 2016). Interestingly, precuneus was the only classic social brain area to be more strongly

15   engaged during normal compared to modulated voice production, supporting previous work

16   suggesting the role of precuneus in self-referential processing (Cabanis et al. 2013).

17   Nevertheless, more work is needed to understand the specific neural and cognitive

18   mechanisms involved in active social behaviour, particularly in interactive settings that allow

19   us to capture the entirety of social interactions, i.e. the action-feedback loop between two

20   interlocutors.

21   **Differentiation of social trait space dimensions in the voice**. Although some evidence

22   points to an integrated, common trait code processing in mPFC and precuneus (Tavares et al.

23   2015; Ma et al. 2016; Van Overwalle et al. 2016), these studies either did not directly

24   compare the two trait dimensions (Ma et al. 2016) or showed differences in the valence of the

25   dimensions (Van Overwalle et al. 2016). Our data suggest that vocal control to express these

1 two components is associated with separable sites of activation and different functional

2 connectivity patterns with vocomotor control networks. Whereas expression of the

3 competence trait led to a stronger engagement of vocomotor control areas in IFG and SMA,

4 affiliative modulations elicited greater activity in a right-dominant network of pSTS/TPJ,

5 SMG, PCC and the left amygdala.

6   The affiliation dimension connotes positive or negative intentions towards others,

7 whereas the competence dimension suggests potency or power to act on such intentions

8 (Fiske et al. 2007). We found that posterior portions of the STS/TPJ were engaged during

9 expression of affiliative information in voices. Processing in the pSTS/TPJ region is

10 particularly dedicated to evaluating others' immediate intentions (Saxe & Powell, 2006; for

11 meta-analytic results see: Molenberghs et al., 2016; Schurz et al., 2014; Van Overwalle,

12 2009) when they are socially significant (Redcay et al. 2010). Given that the affiliation

13 dimension is of high social significance, functional processing in pSTS/TPJ might

14 specifically subserve basic intention encoding of a speaker's own intentions to achieve

15 volitional vocal expression. Keeping with this interpretation, another structure specifically

16 engaged during affiliative vocal modulation was the amygdala, which has frequently been

17 associated with processing behaviourally relevant and salient stimuli (Ewbank et al. 2009).

18 Previous voice research has shown that the amygdala is reactive to affective content in voices

19 (Frühholz and Grandjean 2013; Dricu and Frühholz 2016; Pannese et al. 2016), and is

20 involved in regulating emotional vocal output behaviour (Pichon and Kell 2013; Frühholz et

21 al. 2015). We found that amygdala activation was most pronounced during hostile vocal

22 modulation along the affiliation dimension - this is in line with an affective processing

23 account (LeDoux 2012), but also with previous work showing the amygdala's involvement in

24 encoding stereotypical information along the affiliation dimension in faces (Engell et al.

25 2007; Todorov, Baron, et al. 2008; Todorov, Said, et al. 2008).

1   We observed that intelligent voice modulations were less distinctly perceived than the

2   likeability or hostility expressions. In a perceptual study, a similar network of pSTS/TPJ,

3   SMG, medial portions of STS, PCC, ventral mPFC and amygdala emerged when subjects

4   heard clearly expressed vocal intentions as opposed to ambiguous recordings (Hellbernd and

5   Sammler 2018). Ambiguous expressions engaged SMA, IFG and insula, regions that overlap

6   partly with the competence-related activations we report. The increased engagement of

7   motor-related regions in our study therefore might be due to effects of difficulty in

8   formulating the vocomotor plan to sound "intelligent", leading to perceptually ambiguous

9   expressions. In fact, the fMRI participants judged their own intelligent vocal modulations as

10   less successful than other trait modulations (see supplement S2). Although we accounted for

11   such differences in task difficulty by introducing parametric modulators based on the self-

12   ratings in the statistical models, further studies manipulating social vocal control on both

13   dimensions and reflecting both poles (i.e. decreased as well as increased intelligence) will be

14   needed to differentiate whether the topographical activation differences observed in the

15   current study reflect social or vocomotor processing differences.

16   **Linking social information with motor planning**

17   Our results investigated the interplay of the VMN and SBN in the support of social

18   voice modulation, by examining task-related changes in the functional connectivity of the left

19   IFG. The l-IFG is the central executor providing speech sound maps for voluntary

20   vocalizations (Tourville and Guenther 2011). As a primary input region to the vocomotor

21   network, IFG might serve as a hub for integrating social and vocal information, which is then

22   used to create sound templates for vocomotor translation. We found that SBN regions

23   emerged as being functionally coupled with l-IFG during socially motivated voice production

24   - namely dorsal mPFC, precuneus, right IFG – as well as regions in the basal ganglia.

25   Expressing affiliative information in the voice (speaking in a likeable or hostile voice)

1   revealed additional functional connectivity with right STS. This suggests that during ongoing

2   voice production, social cognitive computations work together with l-IFG to inform motor

3   coordination of the vocal tract. Interestingly, only dorsal mPFC showed significant functional

4   connectivity with l-IFG during social vocal control, but not ventral mPFC. Dorsal mPFC is

5   structurally connected with premotor and somatosensory areas (Öngür et al. 2003), and has

6   been associated with own choice execution (Nicolle et al. 2012) and representing goal

7   oriented social schemata (Krueger et al. 2009). The basal ganglia are part of a subcortical

8   network involved in emotional prosody production (Aziz-Zadeh et al. 2010; Laukka et al.

9   2011; Pichon and Kell 2013; Frühholz et al. 2015; Mitchell et al. 2016) and are thought to

10   have regulatory functional connectivity with the amygdala, motor and auditory cortices

11   during affective vocal control (Pichon and Kell 2013; Frühholz et al. 2015; Klaas et al. 2015).

12   Specifically, ventral and dorsal striatum show distinct roles during emotional versus neutral

13   prosody in motor control planning and executing motor plans, respectively (Pichon and Kell

14   2013). In this previous study, the dorsal striatum was functionally connected to hippocampus,

15   amygdala and motor cortices during angry prosody production. We corroborate these findings

16   showing increased functional connectivity between dorsal striatal regions and l-IFG

17   specifically for ongoing affiliative vocal control and engagement during voice production

18   trials specifically (Go vs. No-Go trials; see supplementary data S3 Table 1). This supports the

19   involvement of the dorsal striatum in volitional vocal control (Laukka et al. 2011), but

20   suggests that this involvement is not exclusively due to affect, but might be a more

21   generalized function for socially relevant volitional vocal modulations.

22   **Limitations.** We acknowledge important differences between the experimental balancing of

23   the two conditions: where both poles of the affiliation dimension were manipulated in the

24   design, only one pole of the competence dimension was implemented. This is because we set

25   out to test specifically the expression of *beneficial* social traits, in the sense that sounding

1 likeable, hostile, and intelligent are helpful in achieving desired interactional outcomes.

2 Whereas trying to come across as intelligent or creating distance or proximity to others may

3 be beneficial, sounding deliberately unintelligent might not naturally lead to valuable

4 outcome. Nevertheless, this imbalance might have impacted comparison between the two

5 dimensions. In addition, although vocal modulations were effective in our study, we cannot

6 account for individual differences in the strategies used to carry out the task. Speakers mainly

7 reported imagining speaking to a known person towards whom they would have liked to

8 express the traits presented, or to whom they had done so in the past. However, this might not

9 be an analogue to how such processes would unfold in novel real-life interactions, and more

10 studies are aiming to provide direct social interaction in fMRI settings to raise ecological

11 validity (Schilbach 2016). Moreover, volitional voice modulation might present a very

12 specific social situation which is rarely practised with such purity in everyday life. We

13 suggest, however, that while voice manipulations may often be the result of spontaneous

14 reactions, the extent of humans' flexible control over vocal expression allows for

15 modulations to be strategically employed in interactions with others.

16   Lastly, we acknowledge that the majority of speakers were female (n=21). Few

17 studies have shown sex-related differences in vocal modulation strategies during courtship,

18 (e.g. Fraccaro et al. 2013; Pisanski et al. 2018). In this study, we were interested in the

19 perceptual effects of vocally expressing socially desirable traits. In fact, relating to the social

20 voice space, McAleer and colleagues (2014) report a consistent pattern in the perception of

21 traits in listeners' ratings of male and female neutral voices (with the exception of perceived

22 attractiveness). This supports the notion that despite the different acoustic modulation

23 strategies that could be present in male and female speaking styles, they nonetheless lead to

24 similar ends in terms of the impressions made on listeners. In fact, to our knowledge only one

25 study has directly investigated volitional vocal control in males and females, suggesting

1 similar efficacy in vocal modulation for most social traits, apart from expressions of

2 confidence (Hughes et al. 2014). Future studies should nevertheless aim to obtain more

3 balanced samples of gender identities to explore potential differences in social voice change

4 and its neural mechanisms.

5 **Conclusions and Outlook**. A number of questions arise from our findings. Our results

6 suggest that social brain areas work together with vocomotor control areas to achieve social

7 vocal control. Although providing some first conclusive results, the underlying neural

8 mechanisms remain unclear, such as which specific social processing functions underlie

9 activation in social brain areas during this type of voice modulation. Moreover, future studies

10 could investigate how individual differences in the efficacy of social voice modulation relate

11 to different levels of social reactivity and mentalizing. For example, are more empathic

12 individuals also better at encoding social information in their voice? Additionally, it remains

13 to be determined how vocal modulation skills arise, as they comprise both social knowledge

14 and fine-tuned motor control. Do we instinctively learn to express social information in the

15 voice in the same way that we learn to speak, via our innate capacity as vocal learners?

16 Lastly, we have introduced an intuitive vocal modulation task that requires targeted social

17 evaluation and forecasting: this could be a candidate for theory of mind tasks of social

18 expression that could be implemented in isolated and dialogic scenarios.

19 This study advances our understanding of the neural mechanisms involved in

20 intentional vocal modulation during encoding of social trait information. We suggest that

21 social vocal control can be exerted to reinforce percepts of traits across the social voice space

22 and is therefore effective in conveying self-referential social intent. Our findings suggest that

23 vocomotor control areas work together with social brain networks to achieve social vocal

24 modulations, thereby extending previous work focussing on affective voice modulation. We

25 suggest that precuneus and mPFC might be engaged in goal- and outcome-oriented self-

1   referential trait processing, while STS activity might relate to intention encoding to achieve

2   volitional social voice change. In sum, this study underlines the importance of the voice as a

3   social behaviour and suggests that vocomotor networks interact with social processing

4   streams to achieve dynamic vocal behaviours, with goal-directed social effects.

5

References

Ackermann H, Hage SR, Ziegler W. 2014. Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. Behav Brain Sci. 72:1–84.

Agnew ZK, Ward L, McGettigan C, Josephs O, Scott SK. 2017. Distinct neural systems for the production of communicative vocal behaviors. bioRxiv. 107441.

Austin JL. 1975. How to do things with words. 2nd ed. Cambridge, MA: Harvard University Press.

Aziz-Zadeh L, Sheng T, Gheytanchi A. 2010. Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. PLoS One. 5:e8759.

Banse R, Scherer KR. 1996. Acoustic profiles in vocal emotion expression. J Pers Soc Psychol. 70:614–636.

Barrett J, Pike GB, Paus T. 2004. The role of the anterior cingulate cortex in pitch variation during sad affect. Eur J Neurosci. 19:458–464.

Bates D, Mächler M, Bolker BM, Walker SC. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw. 67:1–48.

Baus C, McAleer P, Marcoux K, Belin P, Costa A. 2019. Forming social impressions from voices in native and foreign languages. Sci Rep. 9:414.

Belin P, Boehme B, McAleer P. 2017. The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. PLoS One. 12:e0185651.

Belin P, Zatorre RJ. 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport. 14:2105–2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. Nature. 403:309–312.

Belyk M, Brown S. 2016. Pitch underlies activation of the vocal system during affective vocalization. Soc Cogn Affect Neurosci. 11:1078–1088.

Blair KS, Smith BW, Mitchell DG V, Morton J, Vythilingam M, Pessoa L, Fridberg D, Zametkin A, Sturman D, Nelson EE, Drevets WC, Pine DS, Martin A, Blair RJR. 2007. Modulation of emotion by cognition and cognition by emotion. Neuroimage. 35:430–440.

Book A, Methot T, Gauthier N, Hosker-Field A, Forth A, Quinsey V, Molnar D. 2015. The

mask of sanity revisited: Psychopathic traits and affective mimicry. Evol Psychol Sci. 1:91–102.

Brainard DH. 1997. The Psychophysics Toolbox. Spat Vis. 10:433–436.

Brett M, Anton J-LL, Valabregue R, Poline J-B. 2002. Region of interest analysis using an SPM toolbox. Abstract Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Neuroimage. 16:2.

Brown S, Cockett P, Yuan Y. 2019. The neuroscience of Romeo and Juliet: An fMRI study of acting. R Soc Open Sci. 6:181908.

Bzdok D, Schilbach L, Vogeley K, Schneider K, Laird AR, Langner R, Eickhoff SB. 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct Funct. 217:783–796.

Cabanis M, Pyka M, Mehl S, Müller BW, Loos-Jankowiak S, Winterer G, Wölwer W, Musso F, Klingberg S, Rapp AM, Langohr K, Wiedemann G, Herrlich J, Walter H, Wagner M, Schnell K, Vogeley K, Kockler H, Shah NJ, Stöcker T, Thienel R, Pauly K, Krug A, Kircher T. 2013. The precuneus and the insula in self-attributional processes. Cogn Affect Behav Neurosci. 13:330–345.

Cartei V, Cowles W, Banerjee R, Reby D. 2014. Control of voice gender in pre-pubertal children. Br J Dev Psychol. 32:100–106.

Dricu M, Frühholz S. 2016. Perceiving emotional expressions in others: Activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. Neurosci Biobehav Rev.

Engell AD, Haxby J V., Todorov A. 2007. Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. J Cogn Neurosci. 19:1508–1519.

Ewbank MP, Barnard PJ, Croucher CJ, Ramponi C, Calder AJ. 2009. The amygdala response to images with impact. Soc Cogn Affect Neurosci. 4:127–133.

Fiske ST, Cuddy AJC, Glick P. 2007. Universal dimensions of social cognition: warmth and competence. Trends Cogn Sci.

Fiske ST, Cuddy AJC, Glick P, Xu J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. J Pers Soc Psychol. 82:878–902.

Fraccaro PJ, O'Connor JJM, Re DE, Jones BC, DeBruine LM, Feinberg DR. 2013. Faking it: Deliberately altered voice pitch and vocal attractiveness. Anim Behav. 85:127–136.

Frühholz S, Grandjean D. 2013. Amygdala subregions differentially respond and rapidly adapt to threatening voices. Cortex. 49:1394–1403.

Frühholz S, Klaas HS, Patel S, Grandjean D. 2015. Talking in fury: The cortico-subcortical network underlying angry vocalizations. Cereb Cortex. 25:2752–2762.

Golestani N, Pallier C. 2007. Anatomical correlates of foreign speech sound production. Cereb Cortex. 17:929–934.

Gough PM, Nobre AC, Devlin JT. 2005. Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. J Neurosci. 25:8010–8016.

Grice HP. 1975. Logic and conversation. In: Cole P,, Morgan J, editors. Syntax and semantics. New York: Academic Press. p. 41–58.

Hage SR, Nieder A. 2016. Dual Neural Network Model for the Evolution of Speech and Language. Trends Neurosci.

Harris LT, Fiske ST. 2007. Social groups that elicit disgust are differentially processed in mPFC. Soc Cogn Affect Neurosci. 2:45–51.

Harris LT, Todorov A, Fiske ST. 2005. Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind. In: NeuroImage. p. 763–769.

Hellbernd N, Sammler D. 2016. Prosody conveys speaker's intentions: Acoustic cues for speech act perception. J Mem Lang. 88:70–86.

Hellbernd N, Sammler D. 2018. Neural bases of social communicative intentions in speech. Soc Cogn Affect Neurosci. 13:604–615.

Hu C, Di X, Eickhoff SB, Zhang M, Peng K, Guo H, Sui J. 2016. Distinct and common aspects of physical and psychological self-representation in the brain: A meta-analysis of self-bias in facial and self-referential judgements. Neurosci Biobehav Rev.

Hughes SM, Mogilski JK, Harrison MA. 2014. The Perception and Parameters of Intentional Voice Manipulation. J Nonverbal Behav. 38:107–127.

Jiang X, Pell MD. 2017. The sound of confidence and doubt. Speech Commun. 88:106–126.

Klaas HS, Frühholz S, Grandjean D. 2015. Aggressive vocal expressions-an investigation of their underlying neural network. Front Behav Neurosci. 9:121.

Klasen M, von Marschall C, Isman G, Zvyagintsev M, Gur RC, Mathiak K. 2018. Prosody production networks are modulated by sensory cues and social context. Soc Cogn Affect Neurosci. 13:418–429.

Ko SJ, Sadler MS, Galinsky AD. 2014. The Sound of Power: Conveying and Detecting

Hierarchical Rank Through Voice. Psychol Sci. 26:3–14.

Krauss RM, Freyberg R, Morsella E. 2002. Inferring speakers' physical attributes from their voices. J Exp Soc Psychol. 38:618–625.

Kreifelts B, Ethofer T, Shiozawa T, Grodd W, Wildgruber D. 2009. Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. Neuropsychologia. 47:3059–3066.

Krueger F, Barbey AK, Grafman J. 2009. The medial prefrontal cortex mediates social event knowledge. Trends Cogn Sci. 13:103–109.

Laukka P, Åhs F, Furmark T, Fredrikson M. 2011. Neurofunctional correlates of expressed vocal affect in social phobia. Cogn Affect Behav Neurosci. 11:413–425.

LeDoux J. 2012. Rethinking the Emotional Brain. Neuron.

Ma N, Baetens K, Vandekerckhove M, Kestemont J, Fias W, Van Overwalle F. 2014. Traits are represented in the medial prefrontal cortex: An fMRI adaptation study. Soc Cogn Affect Neurosci. 9:1185–1192.

Ma N, Baetens K, Vandekerckhove M, Van der Cruyssen L, Van Overwalle F. 2013. Dissociation of a trait and a valence representation in the mPFC. Soc Cogn Affect Neurosci. 9:1506–1514.

Ma N, Vandekerckhove M, Van Hoeck N, Van Overwalle F. 2012. Distinct recruitment of temporo-parietal junction and medial prefrontal cortex in behavior understanding and trait identification. Soc Neurosci. 7:591–605.

Ma N, Vandekerckhove M, van Overwalle F, Seurinck R, Fias W. 2011. Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas. Soc Neurosci. 6:123–138.

Ma N, Wang S, Yang Q, Feng T, Van Overwalle F. 2016. The neural representation of competence traits: An fMRI study. Sci Rep. 6:39609.

McAleer P, Todorov A, Belin P. 2014. How do you say "hello"? Personality impressions from brief novel voices. PLoS One. 9:e90779.

McGettigan C. 2015. The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication. Front Hum Neurosci. 9:129.

McGettigan C, Eisner F, Agnew ZK, Manly T, Wisbey D, Scott SK. 2013. T'ain't what you

say, it's the way that you say it--left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. J Cogn Neurosci. 25:1875–1886.

Mitchell RLC, Jazdzyk A, Stets M, Kotz SA. 2016. Recruitment of Language-, Emotion- and Speech-Timing Associated Brain Regions for Expressing Emotional Prosody: Investigation of Functional Neuroanatomy with fMRI. Front Hum Neurosci. 10:518.

Molenberghs P, Johnson H, Henry JD, Mattingley JB. 2016. Understanding the minds of others: A neuroimaging meta-analysis. Neurosci Biobehav Rev.

Morningstar M, Dirks MA, Huang S. 2017. Vocal Cues Underlying Youth and Adult Portrayals of Socio-emotional Expressions. J Nonverbal Behav. 41:155–183.

Nicolle A, Klein-Flügge MC, Hunt LT, Vlaev I, Dolan RJ, Behrens TEJ. 2012. An Agent Independent Axis for Executed and Modeled Choice in Medial Prefrontal Cortex. Neuron. 75:1114–1121.

Oleszkiewicz A, Pisanski K, Lachowicz-Tabaczek K, Sorokowska A. 2017. Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. Psychon Bull Rev. 24:856–862.

Öngür D, Ferry AT, Price JL. 2003. Architectonic subdivision of the human orbital and medial prefrontal cortex. J Comp Neurol. 460:425–449.

Pannese A, Grandjean D, Frühholz S. 2016. Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions. Cortex. 85:116–125.

Pavela Banai I, Banai B, Bovan K. 2017. Vocal characteristics of presidential candidates can predict the outcome of actual elections. Evol Hum Behav. 38:309–314.

Peschke C, Ziegler W, Eisenberger J, Baumgaertner A. 2012. Phonological manipulation between speech perception and production activates a parieto-frontal circuit. Neuroimage. 59:788–799.

Pichon S, Kell CA. 2013. Affective and sensorimotor components of emotional prosody generation. J Neurosci. 33:1640–1650.

Pisanski K, Cartei V, McGettigan C, Raine J, Reby D. 2016. Voice Modulation: A Window into the Origins of Human Vocal Control? Trends Cogn Sci.

Pisanski K, Fouquet M, Mathevon N, Reby D. 2016. A longitudinal study of within-individual variation in human voice pitch. J Acoust Soc Am. 140:3397–3397.

Pisanski K, Mora EC, Pisanski A, Reby D, Sorokowski P, Frackowiak T, Feinberg DR. 2016.

Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. Sci Rep. 6:34389.

Pisanski K, Oleszkiewicz A, Plachetka J, Gmiterek M, Reby D. 2018. Voice pitch modulation in human mate choice. Proc R Soc B Biol Sci. 285:20181634.

Ponsot E, Burred JJ, Belin P, Aucouturier JJ. 2018. Cracking the social code of speech prosody using reverse correlation. Proc Natl Acad Sci U S A. 115:3972–3977.

Redcay E. 2008. The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. Neurosci Biobehav Rev.

Redcay E, Dodell-Feder D, Pearrow MJ, Mavros PL, Kleiner M, Gabrieli JDE, Saxe R. 2010. Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. Neuroimage. 50:1639–1647.

Redcay E, Velnoskey KR, Rowe ML. 2016. Perceived communicative intent in gesture and language modulates the superior temporal sulcus. Hum Brain Mapp. 37:3444–3461.

Reiterer SM, Hu X, Sumathi TA, Singh NC. 2013. Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. Front Psychol. 4:782.

Sauter DA, Eisner F, Calder AJ, Scott SK. 2010. Perceptual cues in nonverbal vocal expressions of emotion. Q J Exp Psychol. 63:2251–2272.

Saxe R, Powell LJ. 2006. Specific Brain Regions for One Component of Theory of Mind. Psychol Sci. 17:692–699.

Schall S, Kiebel SJ, Maess B, Von Kriegstein K. 2014. Voice identity recognition: Functional division of the right STS and its behavioral relevance. J Cogn Neurosci. 27:280–291.

Schilbach L. 2016. Towards a second-person neuropsychiatry. Philos Trans R Soc B Biol Sci.

Schroeder J, Epley N. 2015. The Sound of Intellect: Speech Reveals a Thoughtful Mind, Increasing a Job Candidate's Appeal. Psychol Sci. 26:877–891.

Schurz M, Radua J, Aichhorn M, Richlan F, Perner J. 2014. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neurosci Biobehav Rev.

Shultz S, Vouloumanos A, Pelphrey K. 2012. The Superior Temporal Sulcus Differentiates Communicative and Noncommunicative Auditory Signals. J Cogn Neurosci. 24:1224–1232.

Simmonds AJ, Leech R, Iverson P, Wise RJS. 2014. The response of the anterior striatum during adult human vocal learning. J Neurophysiol. 112:792–801.

Simonyan K, Horwitz B. 2011. Laryngeal motor cortex and control of speech in humans.

Neuroscientist. 17:197–208.

Skoog Waller S, Eriksson M. 2016. Vocal Age Disguise: The Role of Fundamental Frequency and Speech Rate and Its Perceived Effects. Front Psychol. 7:1814.

Slotnick SD, Moo LR, Segal JB, Hart J. 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cogn Brain Res. 17:75–82.

Tavares RM, Mendelsohn A, Grossman Y, Williams CH, Shapiro M, Trope Y, Schiller D. 2015. A Map for Social Navigation in the Human Brain. Neuron. 87:231–243.

Thornton XMA, Weaverdyck ME, Tamir XI. 2019. The social brain automatically predicts others' future mental states. J Neurosci. 39:140–148.

Todorov A, Baron SG, Oosterhof NN. 2008. Evaluating face trustworthiness: A model based approach. Soc Cogn Affect Neurosci. 3:119–127.

Todorov A, Mandisodza AN, Goren A, Hall CC. 2005. Psychology: Inferences of competence from faces predict election outcomes. Science (80-). 308:1623–1626.

Todorov A, Said CP, Engell AD, Oosterhof NN. 2008. Understanding evaluation of faces on social dimensions. Trends Cogn Sci. 12:455–460.

Tourville JA, Guenther FH. 2011. The DIVA model: A neural theory of speech acquisition and production. Lang Cogn Process. 26:952–981.

Van Duynslaeger M, Van Overwalle F, Verstraeten E. 2007. Electrophysiological time course and brain areas of spontaneous and intentional trait inferences. Soc Cogn Affect Neurosci. 2:174–188.

Van Overwalle F. 2009. Social cognition and the brain: A meta-analysis. Hum Brain Mapp.

Van Overwalle F, Ma N, Baetens K. 2016. Nice or nerdy? The neural representation of social and competence traits. Soc Neurosci. 11:567–578.

Venables WN, Ripley BD. 2002. Modern Applied Statistics with S. 4th ed. New York: Springer. http://www.stats.ox.ac.uk/pub/MASS4.

Winter K, Spengler S, Bermpohl F, Singer T, Kanske P. 2017. Social cognition in aggressive offenders: Impaired empathy, but intact theory of mind. Sci Rep. 7:670.

Wolpert DM, Doya K, Kawato M. 2003. A unifying computational framework for motor control and social interaction. Philos Trans R Soc B Biol Sci.
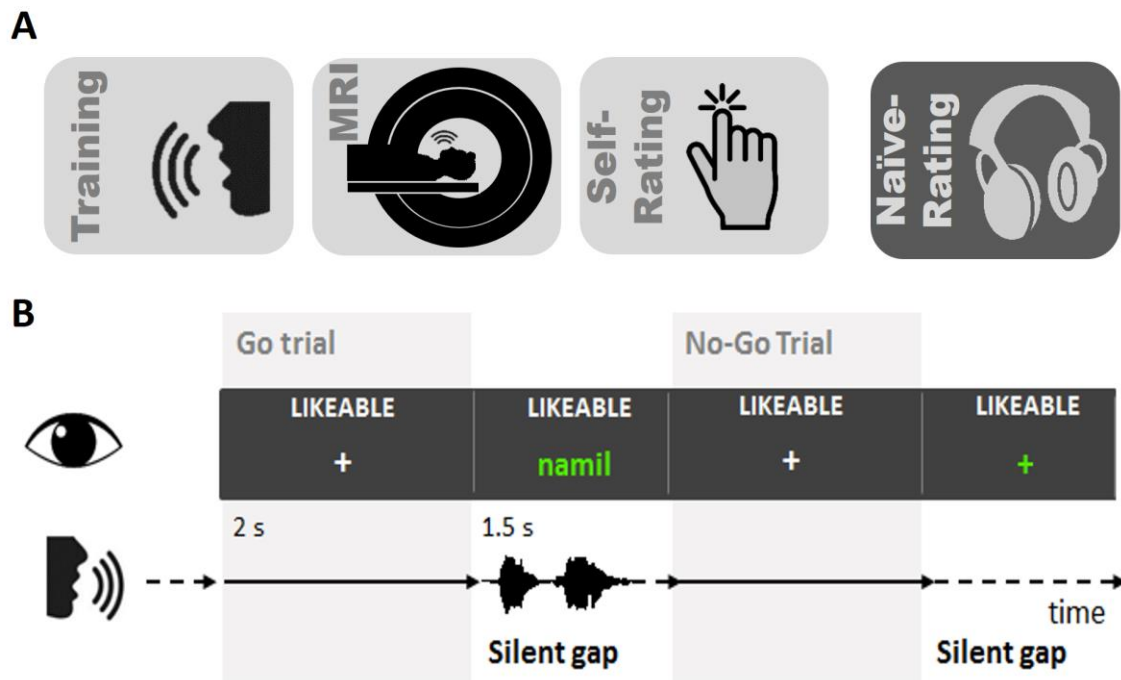
Figure legends and Tables

Figure 1. Experimental procedure and example trial structure of the social vocal control task. **A**. Participants first completed a training session, prior to completing the social vocal control task in the scanner. Post-scanning, each speaker rated all of their own vocal recordings on 7-point Likert-scales (self-rating). Lastly, naïve listeners rated the social vocal control performance of each speaker's best (i.e. self-selected) vocal modulations on all social traits. **B**. During the scanning sessions, participants prepared each social trait expression for 2 s before being shown the relevant exemplar. During Go trials, speech was recorded during the silent inter-scan gap of 1.5 s after the exemplar was presented. During No-Go trials, no exemplar was presented, and the fixation cross remained on the screen but changed its color to green to indicate the duration of the silent gap.
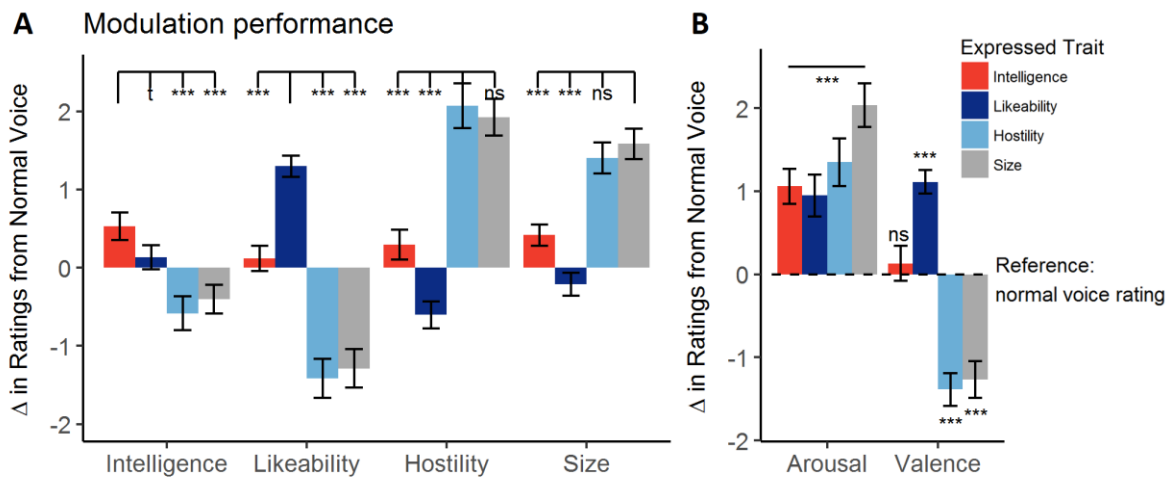
Figure 2. Results from univariate repeated-measures ANOVAs for each rated trait category. **A**. Changes in trait ratings relative to ratings of each speaker's normal voice (Δ-ratings) as a factor of social trait modulation. The models represent the comparison of congruent trait ratings to all incongruent trait ratings. **B**. Changes in intensity ratings on arousal and valence from normal voices as a factor of social trait modulation. ***$p<.001$, ns = not significant. Contrasts were corrected for multiple comparisons using Bonferroni correction. Error bars = standard errors.
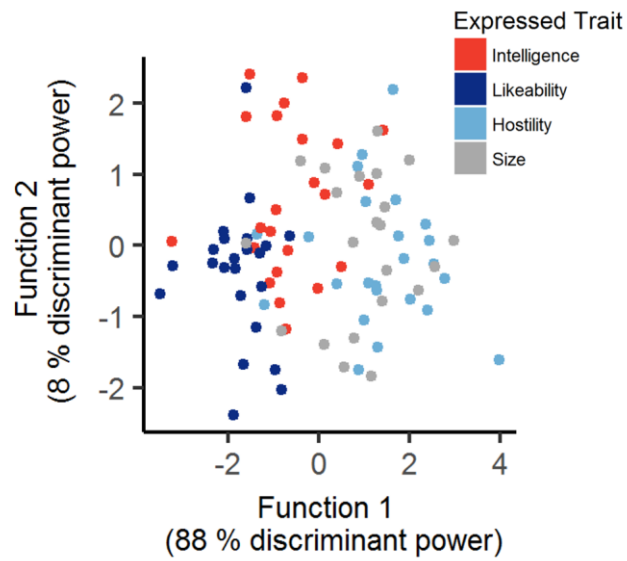
Figure 3. Distribution of linear discriminants of each recording resulting from the first two linear discriminant functions. Colours show the trait expressed by the speakers.
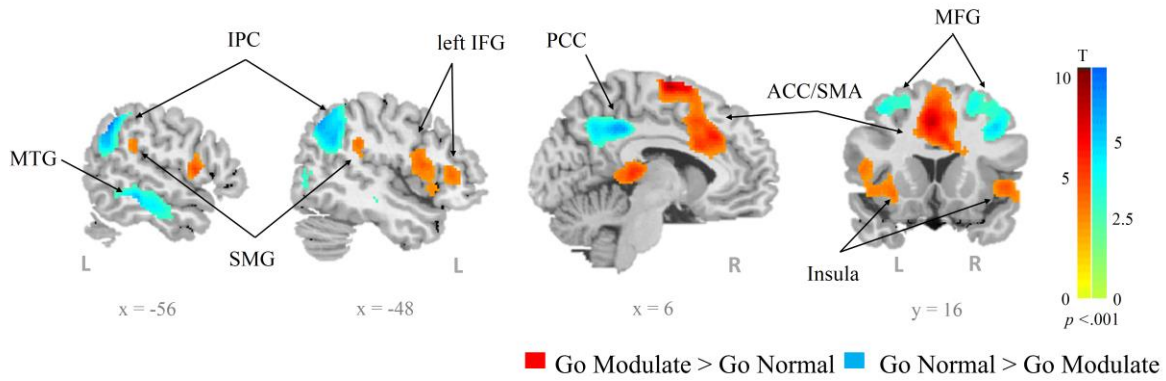
Figure 4. Vocal Modulation Network: Activation maps. The contrast Go Modulation > Go Normal voice (red) evoked changes in activation in ACC, IFG, Insula, SMA, SMG, STG. The contrast Go Normal voice > Go Modulation (blue) showed activation in IPC, MFG, MTG, Precuneus, PCC. ACC = anterior cingulate cortex, IFG = inferior frontal gyrus, IPC = inferior parietal cortex, MFG = middle frontal gyrus, MTG = middle temporal gyrus, SMA = supplementary motor area, SMG = supramarginal gyrus, STG = superior termporal gyrus. L= left, R = right.
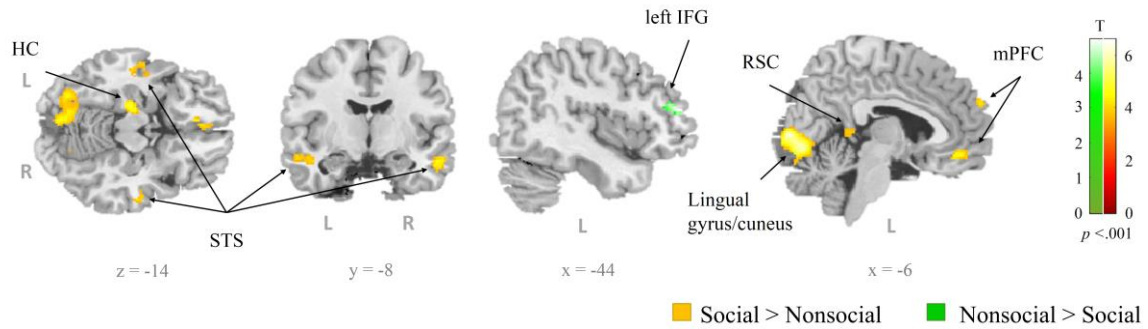
Figure 5. Social voice modulation: Activation maps. The contrast Social > Non-social voice modulation (yellow) activated areas including the bilateral STS, mPFC, left HC cortex, RSC, lingual gyrus, cuneus and precuneus (not depicted). Nonsocial > social voice change (green) lead to changes in activation in left triangular parts of the IFG. IFG = inferior frontal gyrus, mPFC = medial prefrontal cortex, HC = Hippocampus, RSC = retrospenial cortex, STS = superior temporal sulcus. L= left, R = right.
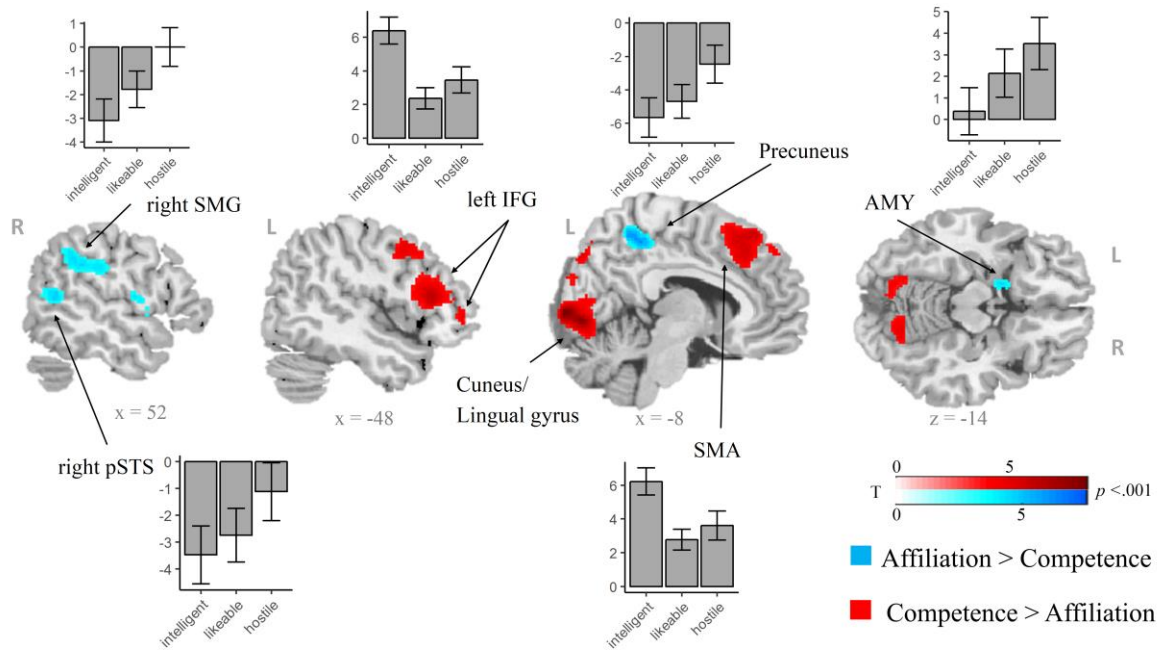
Figure 6. Social trait space modulation: Activation maps and parameter estimates. The contrast Affiliation > Competence (blue) induced changes in activation in clusters including right TPJ, right SMG, precuneus, bilateral AMY. The contrast Competence > Affiliation (red) evoked BOLD changes in left IFG pars opercularis and pars triangularis, ACC, SMA, cuneus and lingual gyrus. Parameter estimates illustrate evoked changes in response to each modulation condition. ACC = anterior cingulate cortex, AMY = amygdala, IFG = inferior frontal gyrus, mPFC = medial prefrontal cortex, HC = Hippocampus, PCC = posterior cingulate cortex, SMA = supplementary motor area, SMG = supramarginal gyrus. Bar plots illustrate parameter estimates (arbitrary units) in the significant cluster per condition compared to rest, error bars = standard errors. L= left, R = right.
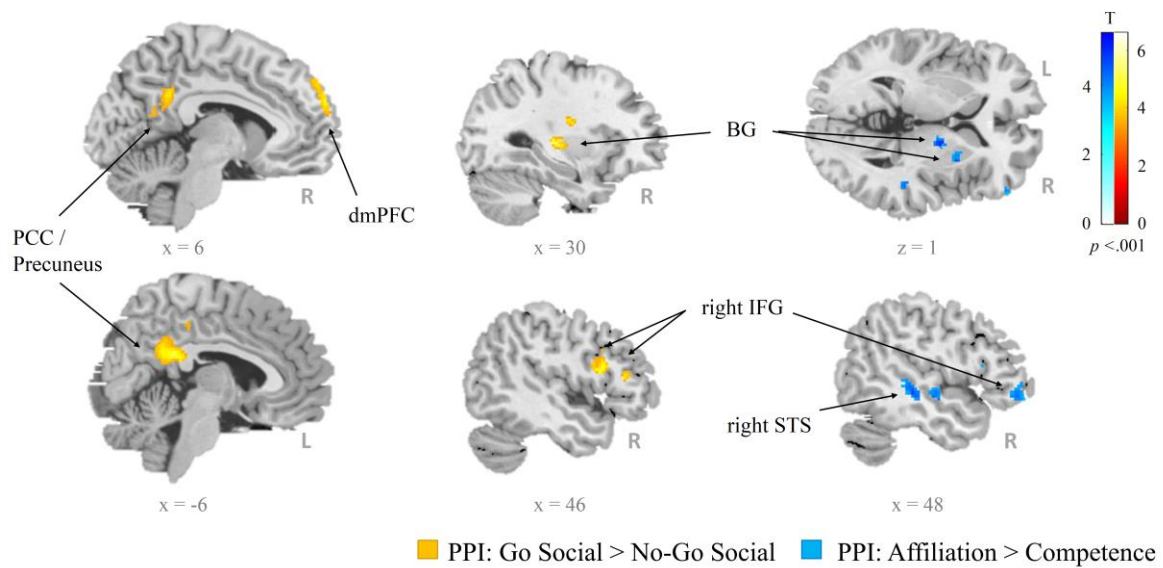
Figure 7. **PPI analysis connectivity maps, seed: left IFG**. Go Social > No-Go Social Voice Modulation (yellow). During social voice modulation left IFG was functionally connected to regions including the medial prefrontal regions, posterior cingulate cortex, BG and right IFG. Affiliation > Competence modulation (blue). Connectivity with the left IFG increased during social voice modulation along the affiliation continuum with right STS and IFG, as well as two clusters in the basal ganglia including (dorsal striatum). Seed: left IFG. BG = basal ganglia, dmPFC = dorso-medial prefrontal cortex, IFG = inferior frontal gyrus, PCC = posterior cingulate cortex. L= left, R = right.

Table 1. Functional activations for the contrasts go modulate > go normal voice and go normal > go modulate.

Contrast	*k*	Region	Hem.	Coordinate			*T*	*Z*
				x	*y*	*z*		
Modulation > normal voice (go trials)	4817	Supplementary motor area	L	0	-4	68	10.55	6.31
		Pre-Supplementary motor area	L	-8	-2	72	9.28	5.93
		Anterior cingulate gyrus	L	-4	16	42	8.58	5.69
	1290	Inferior frontal gyrus, pars triangularis	L	-38	26	2	6.19	4.7
		Inferior frontal gyrus, pars opercularis	L	-56	8	10	5.62	4.41
		Insula	L	-32	18	-4	5.54	4.37
	374	Insula	R	48	12	-2	5.87	4.54
		Superior temporal gyrus	R	56	8	-4	4.82	3.97
			R	62	2	-2	4.7	3.9
	142	Supramarginal gyrus	L	-52	-36	28	4.9	4.01
	865	Corpus callosum, posterior part		0	-28	12	7.5	5.28
			R	12	-36	14	5.23	4.2
			R	4	-22	24	3.79	3.31
Normal voice > Modulation (go trials)	1215	Inferior parietal cortex	L	-54	-52	46	9.13	5.88
			L	-54	-56	32	7.46	5.27
	2530	Inferior parietal cortex	R	54	-54	34	8.92	5.81
			R	56	-48	46	8.06	5.50
			R	48	-60	36	7.99	5.47
	1306	Posterior cingulate cortex	R	6	-42	40	7.57	5.31
			R	2	-58	42	6.36	4.78
	2073	Superior frontal gyrus	R	26	30	46	7.54	5.30
		Middle frontal gyrus	R	40	18	40	6.58	4.89
			R	40	28	46	6.31	4.76
	816	Middle frontal gyrus	L	-34	24	48	6.59	4.89
			L	-38	30	44	6.55	4.87
		Superior frontal gyrus	L	-22	28	50	5.29	4.24
	554	Middle temporal gyrus	L	-56	-34	-10	6.79	4.98
			L	-62	-36	-4	6.4	4.8
			L	-58	-24	-16	5.84	4.53
	238	Middle temporal gyrus	R	62	-22	-14	5.73	4.47
			R	58	-6	-24	4.4	3.71
	108	Cerebellum	L	-38	-74	-34	5.13	4.14
	92	Orbital frontal gyrus	R	42	54	-6	5.10	4.13
		Middle frontal gyrus	R	46	54	4	4.0	3.45

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. $p<.001$ uncorrected, minimal cluster size: 61 voxels.

Table 2. Functional activations for the contrasts social > nonsocial modulation and nonsocial > social modulation.

Contrast	k	Region	Hem.	Coordinate			T	Z
				x	y	z		
Social > Non-social voice change (go trials)	1930	Lingual Gyrus	L	-8	-78	-6	6.61	4.90
				-10	-84	0	6.43	4.81
				0	-88	0	5.77	4.49
	86	Precuneus	L	-20	-48	36	5.02	4.08
	62	Cuneus	L	-20	-92	20	4.17	3.56
				-22	-86	32	3.83	3.33
	105	Hippocampus	L	-22	-20	-14	4.85	3.98
	111	Ventral medial prefrontal cortex	L	-6	44	-12	4.68	3.88
				-12	48	-6	3.86	3.36
	69	Dorsal medial prefrontal cortex	L	-4	62	28	4.45	3.45
	122	Superior temporal sulcus	L	-58	-8	-16	4.17	3.56
				-48	-6	-18	4.15	3.55
				-54	-18	-16	4.02	3.46
	144	Superior temporal sulcus	R	52	-8	-18	4.63	3.85
				58	-14	-14	4.18	3.57
	68	Retrosplenial cortex	L	-6	-40	6	3.9	3.38
			R	2	-46	12	3.81	3.32
Nonsocial > Social voice change (go trials)	65	Inferior frontal gyrus, pars triangularis	L	-44	46	16	4.96	4.05

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. *p*<.001 uncorrected, minimal cluster size: 61 voxels.

Table 3. Functional activations for the contrasts Affiliation > Competence and Competence > Affiliation modulation.

Contrast	*k*	Region	Hem.	Coordinate			*T*	*Z*
				x	*y*	*z*		
Affiliation > Competence	912	Precuneus	L	-12	-42	50	7.47	5.27
			R	2	-42	48	5.08	4.12
		Posterior cingulate gyurs	R	12	-34	44	4.77	3.94
	711	Supramarginal gyrus	R	60	-44	38	5.64	4.42
				58	-36	32	4.97	4.05
				48	-28	32	4.43	3.73
	219	Posterior superior temporal sulcus / Temporo-parietal junction	R	46	-54	14	5.29	4.24
				56	-58	10	5.02	4.08
	97	Amygdala	L	-18	4	-10	5.27	4.23
	74	Precentral gyrus	R	54	-2	10	4.24	3.60
Competence > Affiliation	2526	Cuneus	L	-10	-90	2	8.28	5.58
		Lingual gyrus	L	-8	-80	-2	7.79	5.40
		Calcarine gyrus		0	-84	-2	7.61	5.33
	2026	Superior frontal gyrus	L	-26	20	56	6.61	4.9
				-10	34	54	6.05	4.63
				0	26	46	5.33	4.26
	1218	Inferior frontal gyrus, pars triangularis	L	-48	26	12	6.0	4.61
		Middle frontal gyrus	L	-42	48	0	5.65	4.43
				-34	58	12	5.20	4.19
	67	Inferior frontal gyrus, pars orbitalis	L	-36	30	-6	4.81	3.96

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. *p*<.001 uncorrected, minimal cluster size: 58 voxels.

Table 4. Results of the PPI analysis.

Contrast	k	Region	Hem.	Coordinate			T	Z
Seed: L IFG				x	y	z		
Go > No Go	718	Posterior cingulate cortex	L	-10	-42	22	6.81	4.99
(social modulation)				-4	-50	28	4.61	3.84
			R	4	-46	32	4.52	3.79
	60	Middle cingulate cortex	L	-2	-30	44	4.42	3.72
	103	Dorsal striatum	R	30	-22	4	5.67	4.44
	67	Posterior Insula	R	28	-6	22	5.02	4.09
				32	-14	20	4.49	3.77
	208	Dorsal medial prefrontal cortex	R	6	62	28	4.67	3.87
				4	58	38	4.14	3.54
				20	62	26	3.82	3.33
	283	Inferior frontal gyrus, pars triangularis	R	50	16	20	5.19	4.18
				48	30	14	4.37	3.69
		Inferior frontal gyrus, pars opercularis	R	50	8	22	4.87	4.0
	140	Cerebellum	L	-26	-76	-32	4.77	3.87
				-32	-68	-34	4.2	3.39
				-40	-60	-36	3.82	3.33
No Go > Go	79	Anterior Insula	L	-26	26	6	5.18	4.18
(social modulation)								
Affiliation >	165	Dorsal Striatum	R	16	-2	4	5.49	4.35
Competence			R	28	6	2	4.47	3.75
		Anterior Insula	R	30	24	14	4.33	3.74
	160	Superior temporal sulcus	R	52	-14	4	4.77	3.93
				50	-28	-4	4.68	3.88
				46	-34	0	4.44	3.74
	65	Inferior frontal gyrus, pars triangularis	R	44	22	10	4.69	3.89
				38	28	10	3.70	3.25
	102	Middle frontal gyrus	R	36	12	36	4.66	3.87
				28	14	46	4.24	3.61
	88	Inferior frontal gyrus, pars orbitalis	R	44	44	-10	4.26	3.62
				50	44	2	3.82	3.33
				30	44	-8	3.75	3.28

Note. k, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. $p<.001$ uncorrected, minimal cluster size: 58 voxels.