Understanding the Intentions Behind Man-Made Products Elicits Neural Activity in Areas Dedicated to Mental State Attribution

Nikolaus Steinbeis[1] and Stefan Koelsch[2]

[1]Max-Planck Institute for Human Cognitive and Brain Research, Stephanstr 1a, 04103 Leipzig, Germany and [2]Department of Psychology, Pevensey Building, University of Sussex, Falmer, Brighton BN1 9QH, UK

Trying to understand others is the most pervasive aspect of successful social interaction. To date there is no evidence on whether human products, which signal the workings of a mind in the absence of an explicit agent, also reliably engage neural structures typically associated with mental state attribution. By means of functional magnetic resonance imaging the present study shows that when subjects believe they are listening to a piece of music that was written by a composer (i.e., human product) as opposed to generated by a computer (i.e., nonhuman product), activations in the cortical network typically reported for mental state attribution (anterior medial frontal cortex [aMFC], superior temporal sulcus, and temporal poles) were observed. The activation in the aMFC correlated highly with the extent to which subjects had engaged in attributing the expression of intentions to the composed pieces, as indicated in a postimaging questionnaire. We interpret these findings as indicative of automatic mechanisms, which reflect mental state attribution in the face of any stimulus that potentially signals the working of another mind and conclude that even in the absence of a socially salient stimulus, our environment is still populated by the indirect social signals inherent to human artifacts.

Keywords: fMRI, music, social cognition, theory of mind

Introduction

Being in possession of a theory of mind (also known as the ability to mentalize or adopting an intentional stance) refers to the cognitive capacity to explain and predict other people's behavior by attributing a set of independent mental states (i.e., intentions, beliefs, desires; Frith and Frith 2003). The neural correlates underlying the attribution of mental states have been extensively investigated uncovering an underlying network comprising the anterior medial frontal cortex (aMFC), the superior temporal sulcus (STS)/temporo-parietal junction, as well as the temporal poles (TPs) (Frith and Frith 2003). Paradigms typically entailed the explicit attribution of mental states in narratives (Fletcher et al. 1995; Goel et al. 1995; Gallagher et al. 2000; Vogeley et al. 2001; Ferstl and von Cramon 2003), cartoon stories (Gallagher et al. 2000), and animated shapes (Castelli et al. 2000) or subjects were made to believe they were interacting with a human agent as opposed to a computer (McCabe et al. 2001; Gallagher et al. 2002; Ramnani and Miall 2004), in the latter case reliably activating a core structure of the neural network, namely the aMFC. However, there is a high prevalence of instances in everyday life, where we are confronted with the products of human agents (such as works of art), signaling previously held intentions and performed actions in the explicit absence of the agent him/herself. It is thus unclear whether inanimate objects signal social meaning, such as their creator's intentions and whether we thus implicitly attempt to fathom these.

To address this question, we measured brain responses when subjects listened to what they thought were compositions as opposed to computer-generated pieces of music. Using musical pieces, which were equally plausible to have been composed or generated by a computer, participants were effectively presented with the same stimulus. However in one condition (Composer) were made to believe that the piece had been composed and thus implicitly reflected the expression of a rational agent's intentions, and in another condition (Computer) were made to believe that the pieces had been generated by a computer program and thus, whereas following certain rules, did not reflect the expression of a rational agent's intention. To avoid any memory effects, half the stimuli ($N = 30$) were presented in one condition and the other half were presented in the other condition, which was counterbalanced across subjects. Thus the basic acoustic information was kept identical over all subjects and contrasting the Composer condition against the Computer condition therefore only yielded brain activity specifically related to the participants' attitude taken toward the stimulus. We predicted that should human products, of which music is a most pervasive instance, be processed with regards to the mental states and particularly intentions of those responsible for their inception, then we ought to see a significant increase of activity in brain areas typically associated with attributing mental states, namely the aFMC, the STS and the TP.

Participants were instructed to rate the perceived pleasantness of each piece of music to ensure that sufficient attention was paid to the music. Thus, their task did not focus on the experimental manipulation. In addition, a questionnaire was filled out after the functional imaging session requiring participants to indicate their thoughts during and on the experiment.

Materials and Methods

Subjects

In total 16 subjects (8 males) were investigated, of which only 12 were included in the analysis. The remaining 4 subjects were excluded on the basis of indications given on the postimaging questionnaire, in that they considered it implausible that either the composed pieces had been composed or that the computer pieces had been generated by a computer. The remaining subjects included 7 males and 5 females with a mean age of 24.6 years (age range: 21-31). None of them were professional musicians and some of them had either played an instrument before or were still playing at the time of the experiment. None of them were familiar with the style of music presented in the experiment.

Stimuli

The stimuli were taken from pieces written by composers belonging to the 2nd Vienese School, namely A. Schönberg and A. Webern. This was motivated by the fact that the success of the presently employed paradigm relied on the plausibility of the conditions. The music by Schönberg and Webern in particular is explicitly atonal (dodeca-phonic), thus having no tonal center, which often gives the music a somewhat random character (particularly for the uninformed listener). This apparent randomness predisposes these pieces to be seen as equally likely to be considered as the unintentional clustering of a series of notes, as well as serious composition, intentionally adhering to an underlying system.

We verified this in a rating study prior to the functional magnetic resonance imaging (fMRI) experiment, presenting subjects with a pool of 140 musical excerpts taken from pieces of the 2 composers and asking subjects ($N = 20$) how plausible they thought it, that the piece was composed or computer-generated. No piece was presented to the same participant twice and presentation was counterbalanced across subjects. From the total pool of stimuli, we eventually took 60, which had been considered to be equally highly plausible to have been either composed or computer-generated. This set of 60 pieces with an average duration of 10.6 s was then used for the fMRI experiment.

Excerpts were taken from Schönberg's *Klavierstück,* op. 33a and b, his *Drei Klavierstücke*, op. 11, as well as from Webern's *Variationen für Klavier*, Op. 27, his *Satzstück für Klavier* and the *Klavierstück, im Tempo eines Menuetts*. The pieces were imported from .midi into Cubase (Steinberg Media Technologies GmbH, Hamburg, Germany) and exported using the Grand option and modified with Cool-Edit (sampling rate = 44.1 kHz; 16-bit resolution). Excerpts were chosen from the pieces if they entailed at least one complete phrase and thus constituted individual and therefore credible musical units. On the basis of our own considerations as well as piloting of the stimuli, excerpts were never shorter than 8 and never longer than 13 s. Thus, subjects would have sufficient time to be able to think about the possible intention behind the music and enough stimuli could be presented in the scanning time.

Experimental Procedure

Participants were instructed outside the scanner and told they were going to be presented with musical pieces that had either been composed or generated by a computer. They were told that our interest lay in whether they would perceive the 2 types of music as more or less pleasant and were therefore asked to indicate on a scale of 1–5, where 1 signaled pleasant and 5 unpleasant with neutral at 3, how pleasant or unpleasant they felt each piece to be. Judgments were to be made after each piece of music. The ratings showed that the 2 types of pieces were not perceived differently in terms of valence (see Fig. 1a).

The presentation of stimuli was blocked so that 5 pieces were played consecutively in each condition. Previous piloting studies suggested that this was the ideal design to establish an "agency" context within which the pieces were listened to. The presentation of each piece of music was jittered by 400–2000 ms. There was an interstimulus interval of 6–8 s and an interval between each block of 20–22 s. Blocks were presented in alternate order of condition and participants were cued before each block and piece what kind of piece (composed or computer generated) they were about to hear.

Postimaging Questionnaire

To be able to relate the functional imaging data back to psychological mechanisms occurring while subjects were listening to the music, we also administered a questionnaire on the subject's thoughts during and on the experiment. Items focused on the frequency and degree to which participants had 1) imagined something while listening to the music (Items 1 and 2: *Did you imagine/visualize anything when listening to the compositions/computer pieces? If so, how often and what?*), 2) had thought about the expression of emotions and intentions (Items 3–6: *Did you feel the compositions/computer pieces were trying to express something, such as an intention/emotion? If so, how often and what?*), 3) had daydreamed during the music (Items 7 and 8: *How often did your thoughts drift off and you started daydreaming (e.g. thinking about friends, relationships, study/work?)*, 4) had felt it was plausible that the composed pieces had been composed and that the computer pieces had been computer-generated (Items 9 and 10: *How plausible did it appear to you that the compositions/computer pieces had been composed/generated by a computer?*) and other items on whether subjects, 5) thought the pieces sounded similar or different (Items 11 and 12: *Did the compositions sound similar/different to the computer-generated pieces?*), 6) how pleasant they felt the compositions/computer pieces to have been, and finally 7) how attentively subjects thought they had listened to the music.

On the basis of responses on items 9 and 10 on the perceived plausibility, 4 subjects were excluded from the initial number of 16 scanned subjects in the subsequent statistical analysis. The items on daydreaming and mind-wandering were included, because this has been frequently associated with activity in the aMFC (Mason et al. 2007).

The only difference between the 2 conditions on any of the items was the extent to which participants had thought about intentions being expressed in the music, namely more so for the composed pieces (mean: 3.41) than for the pieces they believed to be computer-generated (mean: 1.91; $P < 0.05$; see Fig. 1b).

Data Acquisition and Analysis

Imaging was performed on a 3T Trio scanner (Siemens, Erlangen, Germany) equipped with a standard bird-cage head coil. A gradient recalled echo-planar imaging (EPI)-sequence was used with time repetition (TR) = 2000 ms and time echo (TE) = 30 ms. A total of 22 axial slices were collected with a slice thickness of 5 mm and an interslice gap of 1 mm. Prior to the functional image acquisition 2 sets of 2-dimensional anatomical images were acquired (T1 Model Driven Fourier Transform [MDEFT] sequence with TR = 1.3 s and TE = 10 ms and an EPI-T1 sequence with the same parameters as the functional run).

Data processing was performed using the software package LIPSIA (Lohmann et al. 2001). Functional data were corrected for motion artifacts and to correct for the temporal offset between slices acquired in one scan, a cubic spline-interpolation was applied. Data were filtered using a temporal highpass filter with a cutoff frequency of 1/128 Hz for baseline correction and a spatial Gaussian filter with 3.768-mm full width at half maximum was applied. Functional slices were aligned with a 3D stereotaxic coordinate reference system (acquired for each subjects individually prior to scanning) by means of a rigid linear registration with 6 degrees of freedom (using 3 rotational and 3 translational parameters acquired during the MDEFT and EPI-T1 sequences). The rotational and translational parameters were sub-sequently transformed by linear scaling to a standard size and the resulting parameters were used to transform the functional slices by using trilinear interpolation (thus, functional slices were aligned with the stereotaxic coordinate system. For the anatomical data, a T_1-weighted, 3D magnetization-prepared rapid gradient-echo sequence was obtained recording a volume data set with 160 slices and 1-mm slice thickness, which was standardized to the Talairach stereotaxic space (Talairach and Tornoux 1988).

Statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations (Worsley and Friston 1995). The design matrix was generated using a synthetic hemodynamic response function. The model equation, including the observed data, the design matrix, and the error term, was convolved with a Gaussian kernel, with a dispersion of 4-s full width at half maximum. Contrast images of the differences between the specified conditions were calculated for each subject. The individual contrast images were then entered into a second-level random effects analysis. Subsequently, t-scores were transformed into Z-scores. On the basis of Monte Carlo Simulations (1000 iterations) with the present brain volume and an individual voxel height threshold of 3.09 ($P < 0.001$, uncorrected), it was determined that a cluster size of 783 mm^3 (29 contiguous voxels) corresponded to an overall image-wise-false-positive rate of 5%. Thus, all activations exceeding this threshold were considered significant at $P < 0.05$, corrected for multiple comparisons. For regions indicated a priori in the experimental hypotheses, we also applied a more liberal threshold of $P < 0.001$, uncorrected.

To correlate some of the ratings given in the questionnaire with activation strength in predicted brain regions, mean beta-values were extracted from the most activated voxel of our hypothesized brain region (in this case aMFC) and determined the 6 adjacent voxels from the mean contrast across participants.

Results

As shown in Figure 1, there were no differences in the perceived emotional valence between pieces played in the Composer condition and the ones played in the Computer condition. However, scores on the questionnaire indicate that participants thought more strongly about the expression of intentions during the Composer condition compared with the Computer condition ($P < 0.05$). There were no further differences between scores for the pieces presented in either condition for any of the other items on the questionnaire.

The fMRI data show that when contrasting the brain activity of the Composer condition against the Computer condition (see Fig. 2 and Table 1), there is an increase in precisely the neuroanatomical network dedicated to mental state attribution, namely the aMFC (–11, 48, 18; $P < 0.05$ corrected), the left STS (mid-portion: –62, –23, 0; $P < 0.05$ corrected and posterior portion: –65, –51, 18; $P = 0.001$ uncorrected) and the right STS (58, –6, –6; $P = 0.001$ uncorrected) as well as the left TP (–50, 7, –21; $P < 0.05$ corrected) and the right TP (37, 15, –30; $P = 0.001$ uncorrected). Notably, the brain activity in the aMFC was correlated highly with the degree to which participants thought that an intention was expressed in the composed pieces of music ($r = 0.76$; $P < 0.01$). There was no increased brain activity when contrasting the Computer condition against the Composer condition.

Discussion

The present study reports the recruitment of a neural network when people engage in the processing of what they believed to
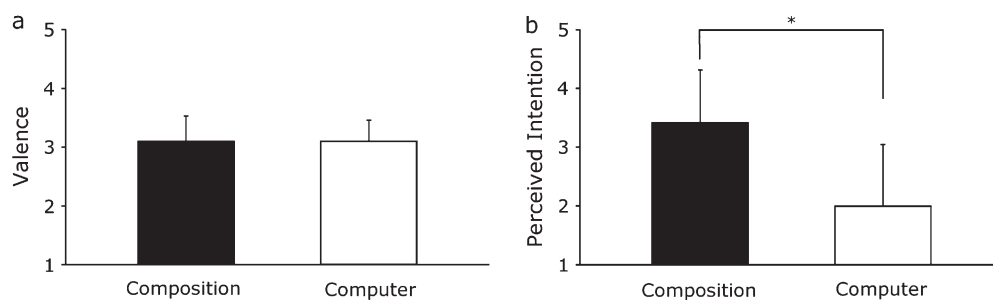
Figure 1. Behavioral ratings. (*a*) Given during the scan on the perceived valence of composed (black) and computer-generated (white) pieces ($P = 0.98$) and (*b*) after the scan on the extent to which participants thought an intention was being expressed, which was higher for composed (black) than computer-generated pieces (white; $P < 0.05$).
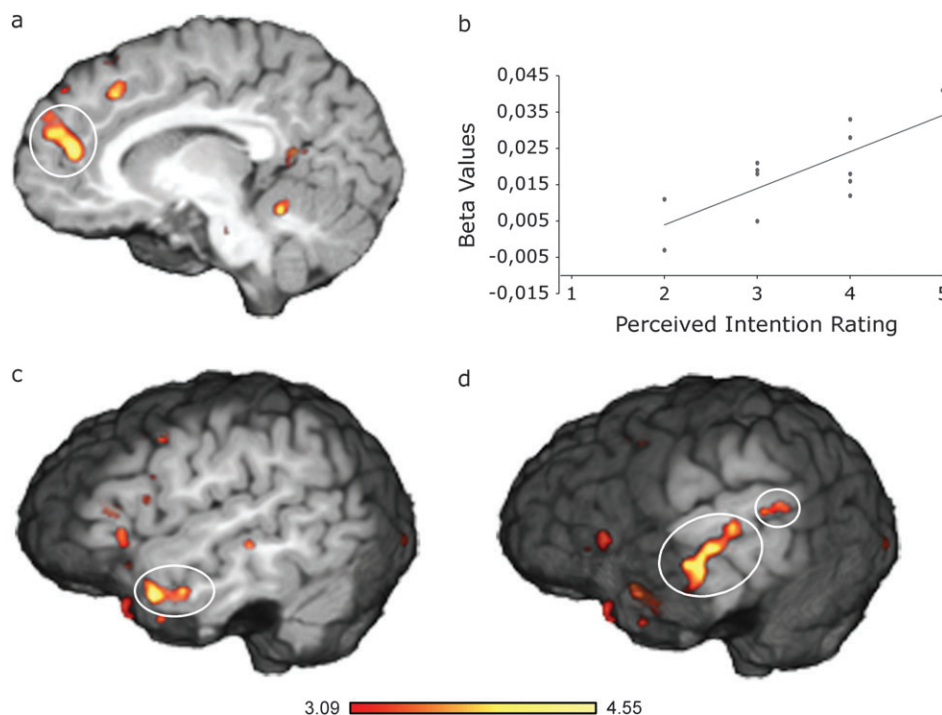
Figure 2. Activations when listening to musical pieces cued as compositions as opposed to computer-generated. (*a*) Increased activation in the aMFC [−11, 48, 18], which (*b*) correlated positively with the extent to which participants thought an intention was being expressed in the compositions ($r = 0.76$; $P < 0.01$). (*c*) Increased activation in the left TP [−50, 7, −21] as well as (*d*) the left STS [−62, −23, 0].

Table 1
Increased activation foci when contrasting compositions against computer-generated pieces

Brain region	Coordinates of peak activation (mm)			Z-score (max)	Extent (mm³)
	x	y	z		
Predicted					
Left aMFC	−11	48	18	4.18*	2079
Left mid-STS	−62	−23	0	4.24*	1647
Left posterior STS	−65	−51	18	3.84	459
Left TP	−50	7	−21	4.27*	1215
Right mid-STS	58	−6	−6	3.56	108
Right TP	37	15	−30	3.56	270
Not predicted					
Left IFG	−41	16	−9	3.87*	945
Left occipital	−32	−95	3	4.26*	972
Right occipital	28	−80	0	3.83*	1188

Note: All activations significant at $P < 0.001$, uncorrected for multiple comparisons; *indicates corrected for multiple comparisons ($P < 0.05$).

be a man-made stimulus as opposed to an artificial stimulus. To our knowledge this is the first study showing that a network engaged in mental state attribution became active when subjects perceive an explicitly non-social stimulus (not containing any first-order sensory information signaling the presence of a human agent). Previous studies using animated shapes (Castelli et al. 2000) both primed subjects to attend to the "feelings and thoughts" of the interacting shapes and the material which elicited the increased activations in the network underlying mental state attribution was more intentional by nature, as indicated by the given ratings. In contrast, in the present study there was no explicit focus on any expressed intentions nor was the material physically different, in actual fact it was equally plausible for it to be random as indicated by the comparable ratings of participants. Thus, our findings clearly demonstrate for the first time, that the attitude alone taken toward a stimulus as social or not is responsible for the increased activations in the neural network underlying mental state attribution.

The functional significance of the individual subcomponents typically reported for mental state attribution has received increased attention recently (Frith and Frith 2003; Amodio and Frith 2006; Saxe 2006). Given that the aMFC was the key region isolated in experiments where participants were made to believe they were interacting with a real human agent as opposed to a computer, it was argued that this region subserves a key component of mental state attribution, that is, to adopt an intentional stance (McCabe et al. 2001; Gallagher et al. 2002; Ramnani and Miall 2004). The coordinates of peak activity reported in these studies strongly resemble the ones reported in the present study (5, 52, 10; McCabe et al. 2001; –10, 50, 30; Gallagher et al. 2002; –8, 56, 24; Ramnani and Miall 2004). The fact that in the present study activity in this region correlated specifically with the degree to which a supposed intention was being expressed lends strong support to the idea that this region reflects the extent to which people think about an intention being expressed. Moreover, recent studies reporting aMFC activity to be specifically modulated by whether participants felt an intention to be communicated or held privately show, disregarding the lateralization, a remarkable overlap with the present peak activation (6, 60, 20; Kampe et al. 2003; 14, 66, 24; Grezes et al. 2004; 0, 54, 12; Walter et al. 2004). Given the correlation with the intention ratings and the overlap with other studies employing an on-line mentalizing

paradigm and the attribution of communicative intentions, the present activation of the aMFC is interpreted as the extent to which participants perceive the piece of music to communicate the (nonspecific) intentions of the composer.

Apart from constituting a key component of the network underlying theory of mind and mental state attribution, the STS has specifically linked to the processing of intentions too (Allison et al. 2000; Castelli et al. 2000; Gallagher et al. 2000; Singer et al. 2004). Similarly to the study by Singer et al. (2004), there was no explicit instruction to focus on the expressed intention of the stimulus and we therefore interpret this structure to automatically process socially relevant events in one's surroundings, something that may have been triggered merely by the cue of an intentional agent's product (i.e., telling participants that they were about to hear composed music).

Within the cortical network underlying mental state attribution, the TPs have been argued to function as a store for relevant personal and semantic knowledge against which the potential meaning of the incoming perceptual information is evaluated (Frith and Frith 2003; Gallagher and Frith 2003). This is supported by recent evidence, that the anterior temporal lobe subserves processing social information providing abstract conceptual knowledge of social behaviors (Zahn et al. 2007). It is possible that participants attempt to match the music and what it is trying to express with what they may have previously heard elsewhere (something the believed computer-generated pieces would automatically be excluded from). Using personally more meaningful music in future studies may be able to shed more light on this yet tentative interpretation.

Our findings show that potentially everything that is man-made is viewed in terms of the expressed intentions of its creator. Thus, our world would appear to be more socially populated than previously believed, as long as an object can be linked to a human agent. Particularly, the meaning of works of art may be derived from the understanding that every note or brush of paint reflects an intentional act, which signals personal relevance to the artist representing a communication between the creator and the perceiver of the artwork. Whereas recent neuroscientific approaches to the perception and appreciation of art and music (Freedberg and Gallese 2007; Molnar-Szakacs and Overy 2006) have stressed the potential involvement of the mirror neuron system in resonating with the artistic expression, the present data would suggest that trying to understand what the artist is attempting to communicate is so far an overruling mechanism determining the understanding of artistic expression.

Notes

Address correspondence to Nikolaus Steinbeis, PhD, Max-Planck Institute for Human Cognitive and Brain Research, Stephanstr 1a, 04103 Leipzig, Germany. Email: steinb@cbs.mpg.de.

References

Allison T, Puce A, McCarthy G. 2000. Social perception from visual cues: role of the STS region. Trends Cogn Sci. 4:267–278.

Amodio D, Frith CD. 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurol. 7:268–277.

Castelli F, Happe F, Frith U, Frith CD. 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage. 12:314–325.

Ferstl EC, von Cramon DY. 2003. What does the frontomedian cortex contribute to language processing: coherence or theory of mind? Neuroimage. 17:1599-1612.

Fletcher PC, Frith CD, Baker SC, Shallice T, Frackowiak RS, Frith CD. 1995. Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. Cognition. 57:109-128.

Freedberg D, Gallese V. 2007. Motion, emotion and empathy in esthetic experience. Trends Cogn Sci. 11:197-203.

Frith U, Frith CD. 2003. Development and neurophysiology of mentalizing. Phil Trans R Soc Lond B. 358:459-473.

Gallagher HL, Frith CD. 2003. Functional imaging of 'theory of mind'. Trends Cogn Sci. 7:77-83.

Gallagher HL, Happe F, Brunswick N, Fletcher P, Frith U, Frith CD. 2000. Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia. 38:11-21.

Gallagher HL, Jack AI, Roepstorff A, Frith CD. 2002. Imaging the intentional stance in a competitive game. Neuroimage. 16:814-821.

Goel V, Grafman J, Sadato N, Hallett M. 1995. Modelling other minds. Neuroreport. 6:1741-1746.

Grezes J, Frith C, Passingham RE. 2004. Brain mechanisms for inferring deceit in the actions of others. J Neurosci. 24:5500-5505.

Kampe K, Frith CD, Frith U. 2003. "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing", regardless of modality. J Neurosci. 23:5258-5263.

Lohmann G, Müller K, Bosch V, Mentzel H, Hessler S, Chen L, Zysset S, von Cramon DY. 2001. Lipsia—a new software system for the evaluation of functional magnetic resonance images of the human brain. Comput Med Imaging Graph. 25:449-457.

Mason MF, Norton MI, Van Horn JD, Wegner DM, Grafton ST, Macrae CN. 2007. Wandering minds. The default network and stimulus-independent thought. Science. 315:393-395.

McCabe K, Houser D, Ryan L, Smith V, Trouard T. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. Proc Natl Acad Sci USA. 98:11832-11835.

Molnar-Szakacs I, Overy K. 2006. Music and mirror neurons: from motion to 'e'motion. Soc Cogn Affect Neurosci. 1:235-241.

Ramnani N, Miall C. 2004. A system in the human brain for predicting the actions of others. Nat Neurosci. 7:85-90.

Saxe R. 2006. Uniquely human social cognition. Curr Opin Neurobiol. 16:235-239.

Singer T, Kiebel SJ, Winston JS, Dolan RJ, Frith CD. 2004. Brain responses to the acquired moral status of faces. Neuron. 41:653-662.

Talairach J, Tornoux P. 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional propositional system—an approach to cerebral imaging. New York: Thieme Medical Publishers.

Vogeley K, Bussfeld P, Newen A, Herrmann S, Happe F, Falkai P, Maier W, Shah NJ, Fink GR, Zilles K. 2001. Mind reading: neural mechanisms of theory of mind and self-perspective. Neuroimage. 14:170-181.

Walter H, Adenzato M, Ciaramidaro A, Enrici I, Pia L, Bara BG. 2004. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. J Cogn Neurosci. 16:1854-1863.

Worsley KJ, Friston KJ. 1995. Analysis of fMRI time-series revisited again. Neuroimage. 2:173-181.

Zahn R, Moll J, Krueger F, Huey ED, Garrido G, Grafman J. 2007. Social concepts are represented in the superior anterior temporal cortex. Proc Natl Acad Sci USA. 104:6430-6435.