

Distinct Neural Systems Involved in Agency and Animacy Detection

Maria Ida Gobbini¹, Claudio Gentili^{2,3}, Emiliano Ricciardi^{2,3},
Claudia Bellucci¹, Pericle Salvini⁴, Cecilia Laschi⁴,
Mario Guazzelli², and Pietro Pietrini²

Abstract

■ We designed an fMRI experiment comparing perception of human faces and robotic faces producing emotional expressions. The purpose of our experiment was to investigate engagement of different parts of the social brain by viewing these animate and inanimate agents. Both human and robotic face expressions evoked activity in face-responsive regions in the fusiform gyrus and STS and in the putative human mirror neuron system. These results suggest that these areas mediate perception of agency, independently of whether the agents are

living or not. By contrast, the human faces evoked stronger activity than did robotic faces in the medial pFC and the anterior temporal cortex—areas associated with the representation of others' mental states (theory of mind), whereas robotic faces evoked stronger activity in areas associated with perception of objects and mechanical movements. Our data demonstrate that the representation of the distinction between animate and inanimate agents involves areas that participate in attribution of mental stance. ■

INTRODUCTION

The human brain has evolved systems that are specialized for understanding others and social interaction (Frith & Frith, 2007; Amodio & Frith, 2006; Saxe, 2006; Mitchell, Banaji, & Macrae, 2005; Haxby, Hoffman, & Gobbini, 2000). Collectively, these systems have been called the “social brain” (Haxby, 2010). One fundamental property of the social brain involves systems for recognizing and understanding animate agents (Wiggett, Pritchard, & Downing, 2009; Kriegeskorte et al., 2008; Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007; Wheatley, Milleville, & Martin, 2007; O'Toole, Jiang, Abdi, & Haxby, 2005; Hanson, Matsuka, & Haxby, 2004; Caramazza & Shelton, 1998). The concepts of “animacy” and “agency,” however, are not co-extensive. Animate entities are living things that can act as agents (although there are extreme examples of sentient animate entities that are incapable of intentional actions). Living things that are not sentient and do not act as agents, such as trees and mushrooms, are not animate. The domain of agents, however, can include inanimate automatons, such as robots, that generate their movements and actions to achieve goals.

The neural representations of animacy and agency may be distinct. Developmentally, the ability to perceive the distinction between agentic and nonagentic movements is evident at an earlier age (Luo, Kaufman, & Baillargeon,

2009; Johnson, 2003; Mandler, 1992) than is the ability to distinguish animate from inanimate agents (Okita & Schwartz, 2006; Lillard, Zeljo, Curenton, & Kaugars, 2000).

We have proposed that the brain areas that have been implicated in the representation of animacy—the occipital and fusiform face areas (OFA and FFA), the extrastriate and fusiform body areas (EBA and FBA), and the posterior STS (pSTS)—may actually be more involved in the representation of agency than in animacy (Gobbini et al., 2007) because these areas are engaged by stimuli that are animate in only a very abstract sense. For example, faces of cartoon figures, which are clearly imaginary, elicit responses in the FFA that are equivalent to responses to pictures of real, human faces (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). Point-light displays that depict biological motion evoke strong responses in the EBA and FBA (Peelen, Wiggett, & Downing, 2006) as well as in the posterior STS (pSTS) and frontal operculum (Gobbini et al., 2007). Even more tellingly, animations of rigid geometric shapes that move in ways that suggest social interactions, the Heider–Simmel animations, evoke strong responses in the lateral fusiform areas, including the FFA (Schultz et al., 2003), the pSTS, and the frontal operculum (Gobbini et al., 2007; Martin & Weisberg, 2003; Castelli, Happé, Frith, & Frith, 2000).

One could argue that stimuli such as cartoon figures, point-light biological motion displays, and the Heider–Simmel social animations are depictions of animate entities. After all, a photograph or a drawing of a face or body is not itself animate but only a depiction of something that

¹University of Bologna, Italy, ²University of Pisa, Italy, ³Fondazione “G. Monasterio” Regione Toscana-C.N.R. Research Area, Pisa, Italy, ⁴Scuola Superiore Sant'Anna, ARTS Lab, Pisa, Italy

is animate. In this vein, one could view the rigid triangles in a Heider–Simmel animation as symbolic markers for real people engaged in a recognizable social interaction. From this perspective, robots provide a more convincing test of the role these brain systems play in the representation of agency versus animacy. Robots do not depict animate agents. Rather, they perform the same actions as animate agents but are definitively and recognizably not animate.

We decided to investigate neural responses to robotic as compared with human facial expressions to distinguish the representation of animacy from the representation of agency. We compared the neural response to social cues, namely facial expressions on the basis of whether the agent producing the expressions is animate (human faces) or inanimate (robotic faces). Our main hypothesis was that both types of agent would activate the face-responsive brain areas that had been associated with animate agents, namely the OFA, FFA, and the pSTS, indicating that these areas play a role in the representation of agency irrespective of whether agents are animate or inanimate. We also hypothesized that the mechanical agent would activate more strongly areas that have been associated with the perception of inanimate objects—the medial fusiform cortex—and with the perception of mechanical movement—the middle temporal gyrus (Beauchamp, Lee, Haxby, & Martin, 2002). Our secondary goal was to explore whether other neural systems that are part of the social brain (Frith & Frith, 2006, 2007; Mitchell et al., 2005), distinguish animate from inanimate agents. In particular, we examined

whether activity in the systems for action understanding—the human mirror neuron system (hMNS; Montgomery, Seeherman, & Haxby, 2009; Ricciardi et al., 2009; Umiltà, 2007; Iacoboni et al., 2005; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996)—and for understanding the mental states of other—the theory of mind (ToM) system (Frith & Frith, 2006; Saxe, 2006; Mitchell, Heatherton, & Macrae, 2002)—differentiated between animate and inanimate agents.

METHODS

Subjects

Twelve healthy, right-handed volunteers with normal or corrected-to-normal vision and no record of neurological or psychiatric illness participated in the study (8 women and 4 men, mean age = 26 ± 4 years). Before the experiment, each subject signed an informed consent approved by the local ethical committee.

Stimuli

Videos of an actor, an actress, and a robot performing the six primary expressions (fear, surprise, disgust, angry, sadness, and happiness; see Figure 1) were used as stimuli while the participants were passively viewing the expressions. Videos with the human faces were produced using a Canon XL1s 3CCD digital video camera and were edited using iMovie (Apple Computer, California). The videos of

Figure 1. Example of the stimuli used during the fMRI experiment (here as still pictures). Still images from the videos for three expressions (anger, fear, and happiness) are shown here. The full set of video stimuli used all six primary expressions (anger, disgust, fear, surprise, happiness, and sadness).



the human faces have been used and validated previously (Montgomery et al., 2009).

The robot used for performing the six primary expressions is the WE-4RII developed by Waseda University, Japan. The WE-4RII possesses several mobile facial elements for emotion expression—eyebrows, lips, eyelids, jaws, and neck—and can even change the color of the cheeks from pale to red (Itoh et al., 2006). The videos with the human faces were taken as a model for the performance of emotions with the robot face. We made six videos, corresponding to the performance of the six primary emotions. Each video lasted 3 sec and was shot according to the following sequence: neutral face, emotion, neutral face. Videos were recorded against a white background and with the robot wearing a black T-shirt (see Figure 1).

The videos of the actor, the actress, and the robot performing emotional expressions were then produced as DVD chapters and burned to a DVD. The DVD was used to present the stimuli in high resolution and projected to a screen. Participants viewed the images via a small mirror placed above their eyes.

Stimuli were blocked on the basis of the actor performing the expressions (female, male, and robot).

Task

Participants passively viewed the expressions. Each block of stimuli was preceded by the cue “view.” Each expression was presented for 3 sec followed by an interval of 7 sec before the following expression appeared. In each block, each primary expression was represented once. The order of the expressions in each block and the order of the blocks were randomized for each participant. Each time series contained two blocks with the expressions performed by the actor, two by the actress, and two by the robot. Blocks were separated by an interval of 15 sec.

Before the fMRI session, participants filled out a multiple choice scale with all six expressions listed to evaluate how accurate they were in interpreting the type of the expressions of the videos. Subjects also were asked to score the intensity of the expressions with a scale from 1 (*low*) to 5 (*very intense*).

Acquisition of Images

Responses to the stimuli were recorded using BOLD contrast fMRI with the acquisition of T2*-weighted gradient-echo planar images on a 1.5-T GE scanner (General Electric, Milwaukee, WI). Each brain volume consisted of 33 contiguous 4-mm-thick axial slices (repetition time = 2.5 sec, echo time = 40 msec, flip angle = 90°, field of view = 24 cm, resolution = 64 × 64).

Three time series were obtained in each fMRI session. Each time series began with 15 sec of rest before the presentation of the stimuli.

High-resolution T1-weighted spoiled gradient recall images (1.2-mm-thick axial slices, repetition time = 5.22 msec,

flip angle = 20°, field of view = 24 cm, resolution = 256 × 256 pixels) were obtained for each subject to provide detailed brain anatomy.

Statistics

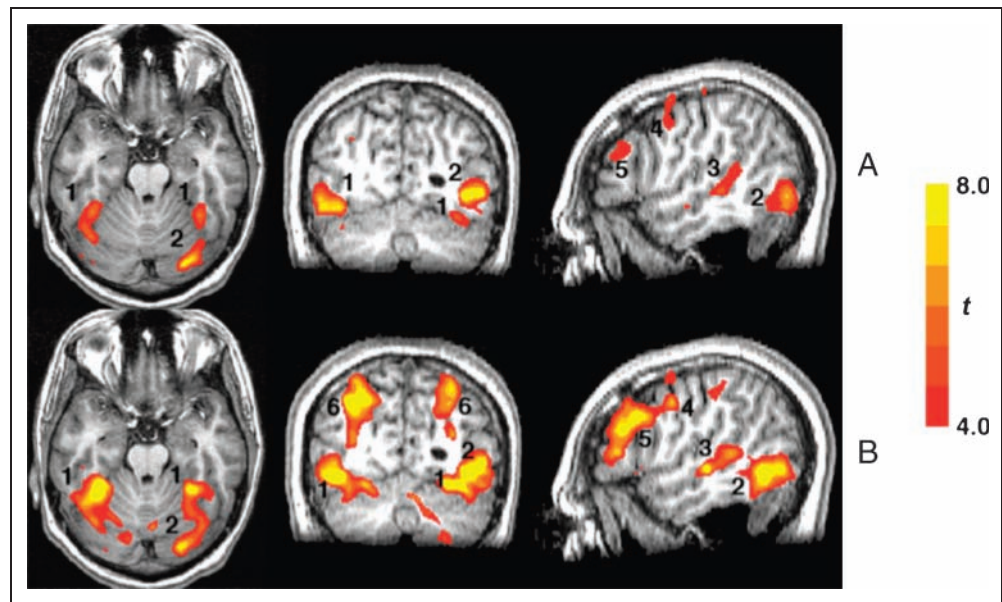
Data analysis was performed with Analysis of Functional Neuroimages (AFNI; <http://afni.nimh.nih.gov>; Cox, 1996). We normalized the time series for each voxel to the mean and smoothed the data spatially (Gaussian kernel 8 mm half-width). Image data were then analyzed with multiple regression. Three regressors of interest (male faces, female faces, and robot faces) were used to model the hemodynamic response with a Gamma function to reflect the time course of the BOLD signal for each expression and

Table 1. Areas Showing Stronger Responses during Passive Viewing of Human Face Expressions as Compared with Rest ($p < .001$)

Area	<i>x</i>	<i>y</i>	<i>z</i>	<i>T</i>	<i>BA</i>
Fusiform gyrus	−20	−85	−11	10.49	19
	38	−47	−19	5.28	37
	−35	−47	−18	5.36	37
Lingual gyrus	24	−88	−4	13.06	18
Middle occipital gyrus	−40	−73	−4	10.15	19
	−44	−74	−10	11.18	19
Superior temporal sulcus	50	−30	−1	5.06	21
IPS	−23	−65	53	3.88	7
Inferior parietal lobule	−51	−38	22	4.11	13
Postcentral gyrus	42	−37	60	6.16	2
	17	−33	62	5.02	4
Precentral gyrus	56	3	37	6.27	6
	−36	−12	65	5.54	6
Premotor area	42	8	28	5.07	9
SMA	−1	−13	67	7.27	6
Inferior frontal gyrus	43	20	14	3.62	45
Middle frontal gyrus	51	31	18	4.7	46
MPFC	3	50	3	3.14	10
Thalamus	20	−31	2	4.47	
Nucleus accumbens	−10	5	−6	4.33	
Amygdala	25	−4	−15	3.41	
Insula	−30	22	7	3.71	
Putamen	−17	9	−1	3.14	
Cerebellum	38	−62	−18	5.34	
	−26	−80	−18	6.53	

IPS = intraparietal sulcus; MPFC = medial prefrontal cortex.

Figure 2. (A) Areas with significantly stronger responses during passive viewing of human face expressions as compared with resting baseline. The threshold for colored regions is $p < .001$, one-tailed. The color scale shows values of $t(df = 11)$. (B) Areas with significantly stronger responses during passive viewing of robotic face expressions as compared with baseline (rest). Axial ($z = -18$), coronal ($y = -67$), and sagittal ($x = 50$) sections are shown. Activated areas: 1 = fusiform gyrus; 2 = middle occipital gyrus; 3 = superior temporal sulcus; 4 = premotor area; 5 = inferior frontal gyrus; 6 = intraparietal sulcus.



a regressor for the cue. Additional regressors of noninterest were used to factor out variance because of overall motion as well as the mean, linear, quadratic, and cubic trends within time series. Thus, the multiple regression model included 3 regressors of interest and 19 regressors of no interest, namely, 1 regressor for the cue condition, 6 regressors for head movements, and 12 regressors that accounted for mean, linear, quadratic, and cubic trends.

The beta coefficients for each regressor of interest were resampled into Talairach space ($1 \times 1 \times 1$ -mm voxels) (Talairach & Tournoux, 1988). We calculated differences between responses to human or robotic faces versus rest and human versus robotic faces for each individual and applied t tests (random effects model) to evaluate group effects. To investigate the areas involved during perception of both human and robotic faces, we performed a conjunction analysis of the comparisons versus rest ($p < .001$, one-tailed, uncorrected, for each comparison, minimum cluster size of 0.5 ml [500 voxels]). To investigate areas that are differentially engaged during perception of human versus robotic faces, we directly compared the responses with these stimuli. Because these differences were small, we used a more liberal threshold ($p < .025$, two-tailed, uncorrected, minimum cluster size of 500 voxels or 0.5 ml), in part to provide a more rigorous test of the absence of differences in the OFA, FFA, and pSTS and in part as an exploratory analysis of differences in other social brain systems. Contrasts in the amygdala were evaluated using a smaller volume threshold (100 voxels or 0.1 ml) because of the small size of this anatomical structure and the previous hypothesis that this structure plays a role in processing emotional expressions.

RESULTS

Behavioral Results

For the subjects in this fMRI study, correct recognition of expressions produced by humans was $84\% \pm 12.5\%$

(mean \pm SD) and by the robotic face $51\% \pm 15\%$ (both greater than chance of 16.7%, $p < .001$). Human expressions were identified correctly more often than robotic expressions ($p < .001$).

Expressions performed by humans were also judged as more intense (3.97 ± 0.5) as compared with the ones performed by the robot (2.74 ± 0.8) ($p < .001$).

Imaging Results

Human Faces versus Rest

Viewing the human faces, as compared with the rest condition, activated the lateral fusiform gyrus bilaterally, the middle occipital gyrus bilaterally, the STS on the right, the precentral gyrus bilaterally, and the postcentral gyrus on the right. At the lower threshold for cluster size, stronger activity also was observed in the right amygdala (for further details on areas of activity, see Table 1; Figure 2).

Robotic Face versus Rest

Viewing the robotic face as compared with rest activated the fusiform gyrus bilaterally, the left middle occipital gyrus, the intraparietal sulcus bilaterally, the middle frontal gyrus bilaterally, the right SMA, and the right STS (for further details on areas of activity, see Table 2; Figure 2).

Overlapping Areas of Activation in the Viewing Condition

Common areas of activity during passive viewing of human faces as compared with rest and of robotic faces as compared with rest were the lateral fusiform gyrus bilaterally, the intraoccipital sulcus, the intraparietal sulcus, the left inferior parietal lobule, the right STS, the right amygdala, the premotor area bilaterally, the middle frontal gyrus bilaterally, the right inferior frontal gyrus, the right SMA, the left

Table 2. Areas Showing Stronger Responses during Passive Viewing of Robotic Face Expressions as Compared with Rest ($p < .001$)

Area	<i>x</i>	<i>y</i>	<i>z</i>	<i>T</i>	<i>BA</i>
Fusiform gyrus	29	-48	-14	18.03	37
	-28	-69	-11	10.21	19
Middle occipital gyrus	-28	-86	3	19.39	18
Intraparietal sulcus	-26	-67	46	6.86	7
	27	-69	45	7.66	7
Paracentral lobule	8	-44	52	5.2	5
Postcentral gyrus	-31	-34	51	8.21	3
Superior temporal sulcus	52	-20	2	10.63	21
Middle frontal gyrus	46	23	21	10.42	46
	-37	25	37	12.77	9
SMA	8	14	44	8.44	6
Cerebellum	-5	-69	-20	7.35	
Thalamus	12	-25	-1	9.17	
Caudate	-14	15	-2	7.43	

cerebellum, the right thalamus, the left putamen, and the left insula (see Tables 1 and 2 and Figure 3).

Direct Contrast of Human versus Robotic Faces in the Viewing Condition

Human faces > robotic faces. Human faces as compared with the robotic faces activated more strongly the medial pFC (MPFC) bilaterally, the left anterior temporal cortex, the right amygdala, and the cuneus. At a lower volume threshold, stronger activity was recorded in the

right TPJ and the right inferior frontal gyrus (Table 3 and Figure 4).

Robotic faces > human faces. The robotic faces activated more strongly the medial fusiform and lingual gyri bilaterally, the middle temporal gyrus bilaterally, the intraparietal sulcus bilaterally, the middle frontal gyrus bilaterally, the left inferior frontal gyrus, the left precentral gyrus, and the right SMA (Table 3 and Figure 4).

DISCUSSION

In this experiment, we investigated the neural systems that participate in the perception of facial expressions produced by humans and by robots. The results showed that the brain areas that are associated with the perception of facial form and movement—the OFA, the FFA, and the pSTS—respond to both human and robotic facial expressions, supporting our hypothesis that these brain areas participate in the representation of agentic forms and actions but do not distinguish between animate and inanimate agents. By contrast, areas that are associated with the representation of the mental states of others—the MPFC, the TPJ, and the anterior temporal cortex—were activated by the human faces but not by the robotic faces, suggesting that these areas do distinguish between animate and inanimate entities. The robotic faces also evoked stronger responses than did human faces in areas that have been associated with the perception of the form of inanimate objects and with mechanical motion—the medial fusiform, the lingual, and the middle temporal gyri—indicating that these areas also distinguish between animate and inanimate entities.

Faces provide a rich source of information that allows us to interact properly with others. The neural systems that mediate this complex function have been extensively investigated. Our model of the neural systems for face perception

Figure 3. Areas with overlapping activations for the comparisons “passive viewing of human faces as compared with baseline” and “passive viewing of robotic faces as compared with baseline.” Axial ($z = -18$), coronal ($y = -67$), and sagittal ($x = 50$) sections are shown. The overlapping areas in yellow in the figure are the fusiform gyrus, the superior temporal sulcus, and the inferior frontal gyrus.

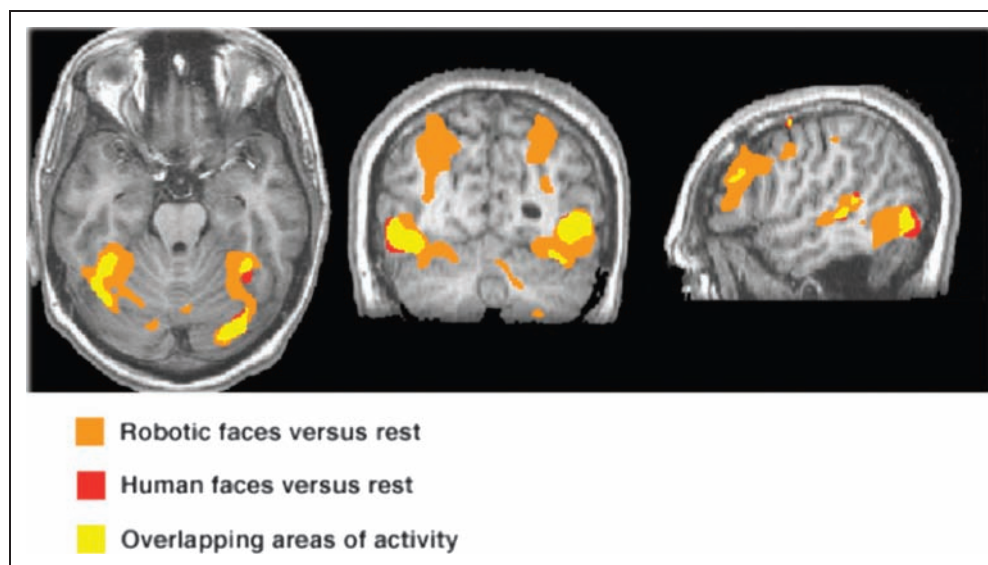


Table 3. Areas Showing Differential Responses to Human versus Robotic Faces during Passive Viewing

<i>Area</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>T</i>	<i>BA</i>
<i>Robots > Humans</i>					
Middle occipital gyrus	31	-79	7	-8.98	19
Fusiform gyrus	27	-76	-16	-4.01	19
	-30	-60	-12	-4.57	19
Parahippocampal gyrus	37	-23	-17	-6.04	36
Middle temporal gyrus	-42	-46	-1	-7.57	37
	56	-46	-6	-5.73	37
IPS	-25	-68	31	-6.43	7
	27	-68	43	-6.13	7
Precentral gyrus	-30	-7	57	-5.11	6
Middle frontal gyrus	47	28	22	-5.19	46
	-41	26	41	-5.69	8
	-36	42	17	-3.64	10
Inferior frontal gyrus (pars orbitalis) ^a	31	25	-13	3.34	47
	43	3	32	-3.57	6
Medial frontal gyrus	-11	4	49	-4.36	24
SMA	11	16	42	-5.24	32
Precentral gyrus	-43	0	31	-4.06	6
Nucleus accumbens	-1	15	-10	3.88	
Insula	41	-3	7	-4.01	
Thalamus	-16	-15	9	-6.93	
Caudate	11	2	21	-3.87	
Cerebellum	-40	-74	-42	-3.54	
	-5	-57	-26	-4.59	
<i>Humans > Robots</i>					
MPFC	6	49	-1	4.36	10
Anterior temporal cortex	-39	-13	-22	3.95	20
TPJ ^b	53	-60	17	3.48	39
Cuneus	11	-88	18	4.08	18
Amygdala ^c	20	-4	-17	3.22	

Areas with volumes <0.5 ml are included that are part of the hMNS or ToM systems (^ainferior frontal gyrus, volume = 0.3 ml; ^bTPJ, volume = 0.14 ml) as well as the amygdala (^cvolume = 0.19 ml).

proposes that these systems are distributed and encompass both visual perceptual area—the “core system”—and areas that play roles in emotion, social cognition, and attention—the “extended system” (Haxby & Gobbini, in press; Gobbini, 2010; Gobbini & Haxby, 2007; Haxby et al., 2000). The results of the current experiment show that viewing human and robotic faces making facial expressions

both engage visual extrastriate areas for face perception—the “core system” (Haxby et al., 2000)—and areas associated with motor representation of facial movement—the putative hMNS. The systems that are differentially engaged by human and robotic facial expressions include those parts of the extended system for face perception that are associated with ToM.

The Core System for Face Perception and Nearby Areas

Looking at human and robotic faces as compared with rest activated similar regions in the extrastriate visual cortex, namely, the lateral fusiform gyri, the middle occipital gyri, and the right STS—the core system for face perception (Haxby et al., 2000). Robotic, but not human, faces additionally evoked stronger responses in the medial fusiform and lingual gyri in the inferior and middle temporal gyri and in the IPS, areas that respond more to objects and mechanical motion than to faces and biological motion (Blakemore et al., 2003; Beauchamp et al., 2002). Our results in these areas are consistent with a report by Mar, Kelley, Heatherton, and Macrae (2007) that compared neural responses with animated depictions and live-action depictions of the same actions.

The lateral fusiform gyrus, including the FFA (Kanwisher, McDermott, & Chun, 1997) and the FBA (Downing, Jiang, Shuman, & Kanwisher, 2001), is involved in the perception of both human and animal faces (Tong et al., 2000) as well as other animate stimuli (Gobbini et al., 2007; Peelen et al., 2006; Grossman & Blake, 2002; Chao, Martin, & Haxby, 1999). Similarly, the right STS is involved in the perception of facial gestures and, more broadly, in the perception of biological motion (Brefczynski-Lewis, Lowitsch, Parsons, Lemieux, & Puce, 2009; Gentili et al., 2008; Engell & Haxby, 2007; Schultz, Friston, O'Doherty, Wolpert, & Frith, 2005; Beauchamp et al., 2002; Grossman & Blake, 2002; Puce, Allison, Bentin, Gore, & McCarthy, 1998; Bonda, Petrides, Ostry, & Evans, 1996). Cartoon faces and schematic faces also activate the fusiform gyrus (Tong et al., 2000); therefore, it is not surprising that robotic faces evoke a neural response in this area. The fusiform gyrus and the STS also are both activated by animations of simple geometrical figures depicting social interactions (Gobbini et al., 2007; Castelli et al., 2000) and by point-light displays depicting biological motion (Gobbini et al., 2007; Peelen et al., 2006; Grossman & Blake, 2002; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001). The fusiform cortex that is activated by social animations with rigid geometrical shapes includes the FFA as defined by a standard functional localizer (Schultz et al., 2003). The range of stimuli that activate these areas, therefore, is not clearly defined by either form or motion. Because of these observations, we proposed that these regions play a more general role in representation of visual stimuli that signify agency (Gobbini et al., 2007). In the current experiment, we tested our hypothesis that the critical factor that engages these regions is the agentic properties

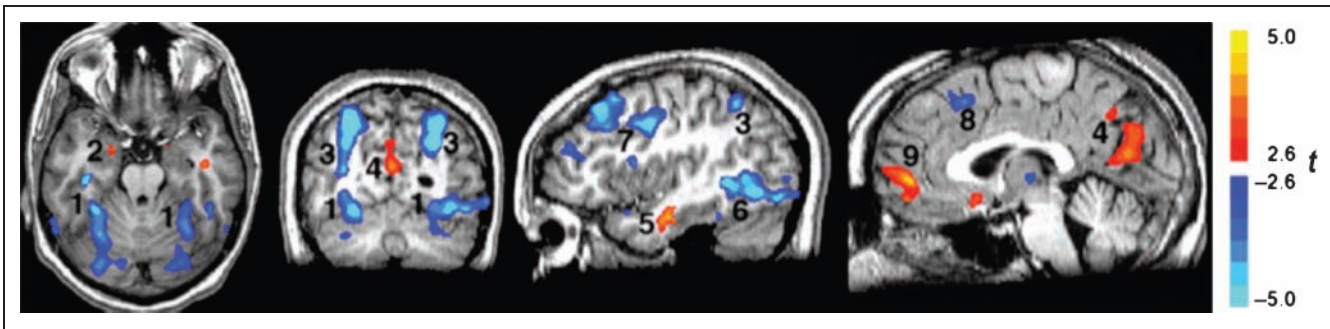


Figure 4. Areas with a differential response to human and robotic face expressions. The threshold for colored regions is $p < .025$, two-tailed. The color scale shows values of $t(df = 11)$. Axial ($z = -14$), coronal ($y = -67$), left sagittal ($x = -40$), and midline sagittal ($x = -2$) sections are shown. Areas showing differential responses: 1 = fusiform gyrus; 2 = amygdala; 3 = intraparietal sulcus; 4 = cuneus; 5 = anterior temporal cortex; 6 = middle and inferior temporal gyrus; 7 = premotor area; 8 = supplementary motor area; 9 = MPFC.

of stimuli rather than their animacy. Robotic stimuli cannot be construed as abstract or degraded depictions of animate entities. They mimic facial expressions but are recognizably and definitively inanimate. Nonetheless, our results show that robots that are designed to produce facial expressions very effectively engage the occipital, fusiform, and superior temporal sulcal cortices for face perception.

The Extended System for Face Perception

Human Mirror Neuron System

Robotic faces as compared with human faces evoked a stronger response in the left premotor cortex, in the right inferior frontal gyrus, and in the right inferior parietal lobule. These regions are considered part of the hMNS (Rizzolatti, Fogassi, & Gallese, 2002; Iacoboni et al., 1999; Decety et al., 1997; Grafton, Fagg, Woods, & Arbib, 1996). Viewing and executing goal-directed and expressive movements evoke activity in the hMNS (Kilner, Neal, Weiskopf, Friston, & Frith, 2009; Montgomery & Haxby, 2008; Buccino et al., 2004; Parsons et al., 1995).

Mentalizing System

Whereas human and robotic faces evoked equivalent responses in the core system areas, human faces evoked stronger responses than did robotic faces in the MPFC, in the left anterior temporal cortex, and in the right amygdala. A trend toward a stronger response was seen also in the right TPJ.

The MPFC, the TPJ, and the anterior temporal cortex are the major components of the mentalizing or ToM system (Amodio & Frith, 2006; Frith & Frith, 2006; Gobbini & Haxby, 2006; Mitchell et al., 2005; Gobbini, Leibenluft, Santiago, & Haxby, 2004; Leibenluft, Gobbini, Harrison, & Haxby, 2004). Mentalizing is the capacity to interpret and predict other people's behavior on the basis of their mental states (Leslie, 1994). Reading false belief stories (Saxe & Kanwisher, 2003; Gallagher et al., 2000), viewing animated geometrical figures that portray social interactions (Gobbini et al., 2007; Martin & Weisberg, 2003; Castelli et al., 2000),

and playing a competitive game against a human partner, as compared with a computer, activate the ToM system (Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004; Gallagher, Jack, Roepstorff, & Frith, 2002; McCabe, Houser, Ryan, Smith, & Trouard, 2001). Perception of eye gaze, an important social cue, also elicits activity in the MPFC (Nummenmaa & Calder, 2009; Calder et al., 2002).

Consistent with our results, others have also found that the MPFC is engaged more strongly by videos with real actors as compared with cartoons (Han, Jiang, Humphreys, Zhou, & Cai, 2005) or computer-generated figures performing the same actions (Mar et al., 2007). Thus, the mentalizing system seems to be sensitive to the animate reality of stimuli. The role of the mentalizing system in the representation of animacy is undoubtedly not equivalent for all animate entities. The concept of animacy includes the "animacy hierarchy" that ranks animate beings on the basis of their level of sentience and awareness—a ranking that affects grammatical forms (Comrie, 1981). Thus, the role of the ToM system is probably greatest for animate beings at the top of the animacy hierarchy.

Inferring Intentions and Differential Activities of the Social Brain

In a recent article, we proposed that the representation of covert mental state—which may predict others' future behavior—and the representation of intentions of perceived, current actions involve distinct neural systems (Gobbini et al., 2007). Intentions that motivate and shape perceived actions are a major component of the representation of those actions in the action understanding system. Mental states that have been associated with the ToM system also include "intentions" that may motivate future behaviors. However, these two types of intentions have distinct representations, including distinct loci of activation in the pSTS and TPJ (Gobbini et al., 2007; Saxe, 2006). The developmental literature also suggests a dissociation between these two levels of understanding others' intentions (Meltzoff, Gopnik, & Repacholi, 1999): The representations of others' intentions that are implied by perceived actions develop before the representation of the covert mental states of

others that predict future actions. The first level of intention understanding—involving a visual or a visuomotor analysis of the scene—might be mediated by the action understanding system (the pSTS and the hMNS) (Gazzola, Rizzolatti, Wicker, & Keysers, 2007), whereas the second level of intention understanding—involving a more abstract and descriptive content—might be mediated by the ToM system (Gobbini et al., 2007; Walter et al., 2004). Both human and robotic faces activated the areas associated with the representation of the intentions associated with perceived actions, but only human faces activated the areas associated with the representation of covert intentions. The robotic faces actually evoked a stronger response in parietal and premotor areas, which may reflect the greater difficulty the subjects had interpreting the robotic facial expressions. Perception of robotic facial expressions that are more realistic and easily interpreted may eliminate this difference in areas associated with attention and the representation of intentions associated with actions.

Robotics and the Social Brain

Building robots that can interact and communicate efficiently with humans is a major goal in robotics (Breazeal, 2005; Cañamero, 2005). The evolution from industrial to service applications (Engelberger, 1989) has been characterized by closer interactions between human beings and robots. Robots are now designed not only as instruments but also as companions and collaborators. Thus, the integration of social capabilities, such as the expression of emotion, in the design of the next generation of robots will be crucial for improving social and collaborative interactions with human beings (Norman, 2004). Efficient engagement of the social brain by robots may facilitate effective human–robot interaction. A robot that possesses social skills can take advantage of the innate human ability to understand and engage in social interactions (Breazeal, 2005).

Acknowledgments

The authors thank Atsuo Takanishi for making his laboratory available for the preparation of the robotic stimuli. They also thank Massimiliano Zecca and all the students of the RoboCasa Laboratory, Waseda University, for the invaluable help with stimulus preparation. They also thank Jim Haxby, Piergiorgio Battistelli, and Carlo Cipolli for helpful discussion. Moreover, they thank Cassa di Risparmio di Pisa (grant eSMILER to P.P.) and the National Science Foundation (grant no. 0830136) for their financial support of this project.

Reprint requests should be sent to Maria Ida Gobbini, Dipartimento di Psicologia, Università di Bologna, Viale C. Berti-Pichat, 5, 40126 Bologna, Italy, or via e-mail: mariaida.gobbini@unibo.it.

REFERENCES

- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7, 268–277.
- Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, 34, 149–159.
- Blakemore, S. J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., & Decety, J. (2003). The detection of contingency and animacy from simple animations in the human brain. *Cerebral Cortex*, 13, 837–844.
- Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, 16, 3737–3744.
- Breazeal, C. L. (2005). *Designing sociable robots*. Cambridge, MA: MIT Press.
- Brefczynski-Lewis, J., Lowitzsch, S., Parsons, M., Lemieux, S., & Puce, A. (2009). Audiovisual non-verbal dynamic faces elicit converging fMRI and ERP responses. *Brain Topography*, 21, 193–206.
- Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., et al. (2004). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron*, 42, 323–334.
- Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., et al. (2002). Reading the mind from eye gaze. *Neuropsychologia*, 40, 1129–1138.
- Cañamero, L. (2005). Emotion understanding from the perspective of autonomous robots research. *Neural Networks*, 18, 445–455.
- Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, 10, 1–34.
- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12, 314–325.
- Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for or faces? *NeuroReport*, 10, 2945–2950.
- Comrie, B. (1981). *Language universals and linguistic typology*. Oxford: Basil Blackwell.
- Connolly, A. C., Abdi, H., & Haxby, J. V. (submitted). Taxonomic organization of animal classes in distributed neural code in ventral temporal cortex.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29, 162–173.
- Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain*, 120, 1763–1777.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2470–2473.
- Engelberger, J. F. (1989). *Robotics in service*. Cambridge, MA: MIT Press.
- Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, 45, 3234–3241.
- Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research*, 1079, 36–46.
- Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, 17, R724–R732.
- Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (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, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage*, 16, 814–821.

- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593–609.
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *Neuroimage*, 35, 1674–1684.
- Gentili, C., Gobbini, M. I., Ricciardi, E., Vanello, N., Pietrini, P., Haxby, J. V., et al. (2008). Differential modulation of neural activity throughout the distributed neural system for face perception in patients with Social Phobia and healthy subjects. *Brain Research Bulletin*, 77, 286–292.
- Gobbini, M. I. (2010). Distributed process for retrieval of person knowledge. In A. Todorov, S. T. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 40–53). New York: Oxford University Press.
- Gobbini, M. I., & Haxby, J. V. (2006). Neural response to the visual familiarity of faces. *Brain Research Bulletin*, 71, 76–82.
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45, 32–41.
- Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, 19, 1803–1814.
- Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage*, 22, 1628–1635.
- Grafton, S. T., Fagg, A. H., Woods, R. P., & Arbib, M. A. (1996). Functional anatomy of pointing and grasping in humans. *Cerebral Cortex*, 6, 226–237.
- Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, 35, 1167–1175.
- Han, S., Jiang, Y., Humphreys, G. W., Zhou, T., & Cai, P. (2005). Distinct neural substrates for the perception of real and virtual visual worlds. *Neuroimage*, 24, 928–935.
- Hanson, S. J., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a “face” area? *Neuroimage*, 23, 156–166.
- Haxby, J. V. (2010). Social neuroscience and the representation of others: Commentary. In A. Todorov, S. T. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 77–81). New York: Oxford University Press.
- Haxby, J. V., & Gobbini, M. I. (in press). Distributed neural systems for face perception. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *Handbook of face perception*. Oxford: Oxford University Press.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4, 223–233.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one’s own mirror neuron system. *PLoS Biology*, 3, 529–535.
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H. L., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286, 2526–2528.
- Itoh, K., Miwa, H., Nukariya, Y., Zecca, M., Takanobu, H., Roccella, S., et al. (2006). *Mechanisms and functions for a humanoid robot to express human-like emotions*. Proceedings of the 2006 IEEE International Conference on Robotics and Automation, Orlando, Florida, USA.
- Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 358, 549–559.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311.
- Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience*, 29, 10153–10159.
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60, 1126–1141.
- Leibenluft, E., Gobbini, M. I., Harrison, T., & Haxby, J. V. (2004). Mothers’ neural activation in response to pictures of their children and other children. *Biological Psychiatry*, 56, 225–232.
- Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, 50, 211–238.
- Lillard, A. S., Zeljo, A., Curenton, S., & Kaugars, A. S. (2000). Children’s understanding of the animacy constraint on pretense. *Merrill-Palmer Quarterly*, 46, 21–44.
- Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants’ reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, 58, 441–486.
- Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, 4, 587–604.
- Mar, R. A., Kelley, W. M., Heatherton, T. F., & Macrae, C. N. (2007). Detecting agency from the biological motion of veridical vs animated agents. *Social Cognitive and Affective Neuroscience*, 2, 199–205.
- Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, 20, 575–587.
- McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences, U.S.A.*, 98, 11832–11835.
- Meltzoff, A. N., Gopnik, A., & Repacholi, B. M. (1999). Toddlers’ understanding of intention, desires and emotions: Exploration of the he dark ages. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention* (pp. 17–41). Mahwah, NJ: Erlbaum.
- Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage*, 28, 757–762.
- Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences, U.S.A.*, 99, 15238–15243.
- Montgomery, K. J., & Haxby, J. V. (2008). Mirror neuron system differentially activated by facial expressions and social hand gestures: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 20, 1866–1877.
- Montgomery, K. J., Seherman, K. R., & Haxby, J. V. (2009). The well-tempered social brain. *Psychological Science*, 20, 211–213.
- Norman, D. A. (2004). *Emotional design. Why we love (or hate) everyday things*. New York: Basic Books.
- Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, 13, 135–143.
- Okita, S. Y., & Schwartz, D. L. (2006). Young children’s understanding of animacy and entertainment robots. *International Journal of Human Robotics*, 3, 392–412.
- O’Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces

- in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17, 580–590.
- Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., et al. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature*, 375, 54–58.
- Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron*, 49, 815–822.
- Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, 18, 2188–2199.
- Ricciardi, E., Bonino, D., Sani, L., Vecchi, T., Guazzelli, M., Haxby, J. V., et al. (2009). Do we really need vision? How blind people “see” the actions of others. *Journal of Neuroscience*, 29, 9719–9724.
- Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage*, 22, 1694–1703.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology*, 12, 149–154.
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, 16, 235–239.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind.” *Neuroimage*, 19, 1835–1842.
- Schultz, J., Friston, K. J., O’Doherty, J., Wolpert, D. M., & Frith, C. D. (2005). Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron*, 45, 625–635.
- Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., et al. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 358, 415–427.
- Talairach, J., & Tournoux, P. (1988). *Co-planar atlas of the human brain*. New York: Thieme.
- Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology*, 17, 257–279.
- Umiltà, C. (2007). Consciousness and control of action. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 327–352). Cambridge, UK: Cambridge University Press.
- Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences, U.S.A.*, 98, 11656–11661.
- Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, 16, 1854–1863.
- Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system. *Psychological Science*, 18, 469–474.
- Wiggett, A. J., Pritchard, I. C., & Downing, P. E. (2009). Animate and inanimate objects in human visual cortex: Evidence for task-independent category effects. *Neuropsychologia*, 47, 3111–3117.