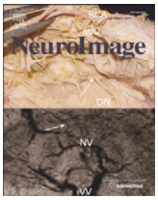
NeuroImage 59 (2012) 3050–3059

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



An integrative model of the neural systems supporting the comprehension of observed emotional behavior

Robert P. Spunt [*], Matthew D. Lieberman

University of California Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 5 August 2011
Revised 7 September 2011
Accepted 4 October 2011
Available online 13 October 2011

Keywords:
Social cognition
Emotion understanding
Mirror systems
Mentalizing system
fMRI

ABSTRACT

Understanding others' emotions requires both the identification of overt behaviors ("smiling") and the attribution of behaviors to a cause ("friendly disposition"). Previous research suggests that whereas emotion identification depends on a cortical *mirror system* that enables the embodiment of observed motor behavior within one's own motor system, causal attribution for emotion depends on a separate cortical *mentalizing system*, so-named because its function is associated with mental state representation. We used fMRI to test an Identification–Attribution model of mirror and mentalizing system contributions to the comprehension of emotional behavior. Normal volunteers watched a set of ecologically valid videos of human emotional displays. During each viewing, volunteers either identified an emotion-relevant motor behavior (*explicit identification*) or inferred a plausible social cause (*explicit attribution*). These explicit identification and attribution goals strongly distinguished activity in the mirror and mentalizing systems, respectively. However, frontal mirror areas, though preferentially engaged by the identification goal, nevertheless exhibited activation when observers possessed the attribution goal. One of these areas—right posterior inferior frontal gyrus—demonstrated effective connectivity with areas of the mentalizing system during attributional processing. These results support an integrative model of the neural systems supporting the comprehension of emotional behavior, where the mirror system helps facilitate the rapid identification of emotional expressions that then serve as inputs to attributional processing in the mentalizing system.

Introduction

People are capable of understanding the inner emotional life of another by simply looking at the outer expression on their face. This capacity is a vital part of normal social cognition, a fact that is apparent from the debilitating nature of psychopathologies that feature impairments in emotion understanding, such as autism spectrum disorder (Blair, 2005). Research investigating the neural bases of emotion understanding and social cognition more broadly has demonstrated the involvement of two anatomically and functionally dissociable brain systems: the so-called *mirror system*, the core of which involves matching observed motor acts to corresponding motor representations in the observer (Iacoboni, 2009; Niedenthal et al., 2010), and the so-called *mentalizing system*, which bears its name because its activity is reliably associated with the representation of the mental states of others (Frith and Frith, 2006; Mitchell, 2009; Saxe, 2006). Though there is general consensus that both systems contribute to emotion understanding (Bastiaansen et al., 2009; Olsson and Ochsner, 2008; Shamay-Tsoory, 2011; Zaki and Ochsner, 2011), the nature of their contribution, as well as the relationship between the two systems, remains unclear. In the present study, we used functional magnetic resonance imaging (fMRI) to test an integrative model of mirror and mentalizing system contributions to understanding emotional facial expressions.

How does the brain understand the emotional states of other brains? One account focuses on a mechanism termed *embodied simulation* (Bastiaansen et al., 2009; Gallese, 2007; Niedenthal et al., 2010), which is based on the fact that covert emotional states (e.g., happiness) are associated with overt motor behaviors (e.g., smiling). Given this, observers can simulate the unobservable emotional state of another by embodying their observable motor state. The existence of simulative processes in emotion perception is supported by the well-documented observation that individuals spontaneously and rapidly mimic other people's facial expressions (Dimberg et al., 2000), and there is evidence that such mimicry is causally related to emotion identification (Neal and Chartrand, 2011). The existence of spontaneous facial mimicry suggests a perception-action matching mechanism in the brain that allows the direct mapping of perceived facial expressions onto the observer's ability to produce the same or similar expressions (Preston and De Waal, 2001). In humans, this mechanism is putatively based in a *mirror system* for observed motor actions, which includes posterior inferior frontal gyrus (pIFG), dorsal premotor cortex (dPMC), and rostral inferior parietal lobule (rIPL). These regions are reliably active during the perception

[*] Corresponding author at: 1285 Franz Hall, Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095-1563, USA.
E-mail address: bobspunt@gmail.com (R.P. Spunt).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.10.005

of motor actions, including actions of the face (e.g., Buccino et al., 2001), and during the execution of motor actions, including the imitation of facial expressions (e.g., Carr et al., 2003). Moreover, these regions are believed homologous to areas of the brain in which single cells with perception-action matching ("mirror") properties have been extensively studied in the macaque (di Pellegrino et al., 1992; Fogassi et al., 2005; Tkach et al., 2007) and more recently in humans (Mukamel et al., 2010).

It has been suggested that the mirror system provides a basis not just for emotion understanding, but also for all domains of social cognition (Gallese et al., 2004). Similarly, a dysfunctional mirror system has been proposed as the basis of the severe social deficits present in psychopathologies such as autism spectrum disorder (Oberman and Ramachandran, 2007). However, numerous theoretically and empirically based critiques of the mirror system account of social cognition have been advanced (Baird et al., 2011; Decety, 2010; Heyes, 2010; Hickok, 2009; Jacob, 2008; Jacob and Jeannerod, 2005; Saxe, 2005; Southgate and Hamilton, 2008). One particularly powerful critique is based on the empirical fact that neuroimaging studies which explicitly ask participants to make judgments regarding the internal states of others, such as their beliefs (Saxe and Kanwisher, 2003), preferences (Mitchell et al., 2006) or emotional state (Budell et al., 2010; Ochsner et al., 2004), reliably recruits a different set of cortical brain regions collectively known as the *mentalizing system*. This system includes dorsomedial and ventromedial prefrontal (dm/vmPFC) cortices, posterior cingulate cortex/precuneus (PCC/PC), temporoparietal parietal junction (TPJ), the posterior superior temporal sulcus (pSTS), and the anterior temporal cortex (aTC) (Frith and Frith, 2006; Mitchell, 2009; Saxe, 2006). The mentalizing system is anatomically independent from the mirror system, a fact that severely undermines the notion that the mirror system is the primary basis for emotion understanding (Keysers and Gazzola, 2007; Olsson and Ochsner, 2008).

In addition to being anatomically independent, there is a great deal of evidence suggesting that the two systems are either functionally independent or even competitive. A recent meta-analysis of over 220 neuroimaging studies of social cognition found that the two systems are rarely concurrently active and concluded that neither system aids or subserves the other (Van Overwalle and Baetens, 2009). This meta-analysis is generally consistent with studies demonstrating that during emotion perception, the two systems appear to process distinct categories of social information, with the mirror system engaged by nonverbal, motor features and the mentalizing system engaged by either contextualizing verbal information (cf. Waytz and Mitchell, 2011; Zaki et al., 2010) or the explicit evaluation of another's emotional state (Budell et al., 2010). Finally, there is evidence suggesting that under some conditions the two systems may actually interfere with one another. The two systems demonstrate anti-correlated activity when individuals are at rest (Fox et al., 2005), and other work suggests that areas of the mentalizing system may operate to inhibit the tendency to imitate another's action, a function putatively based in the mirror system (Spengler et al., 2009).

Contrary to evidence that the two systems are either independent or competitive, a handful of studies suggest they may cooperate during social cognition (cf. Zaki and Ochsner, 2011). Several studies have shown that the two systems exhibit concurrent activation during the observation of complex social stimuli (Brass et al., 2007; Iacoboni et al., 2004), especially when observers are explicitly induced to make judgments regarding the target's internal state (Spunt et al., 2011). Another study demonstrated that activity in both systems positively predict the accuracy of observers' ratings of another person's emotional state (Zaki et al., 2009). Finally, one study found that areas of the two systems demonstrate effective connectivity when participants' are asked to estimate the opinions of another person (Lombardo et al., 2010). These studies suggest the possibility that these two systems may work in tandem to enable emotion understanding, but no study to date has explicitly tested this possibility.

In the present study, we explicitly test an integrative model of mirror and mentalizing system contributions to emotion understanding (Fig. 1A). Our model is derived from classic work in social psychology on topic of social causal attribution (Gilbert, 1998; Kelley, 1973), which states that attributional inferences depend on a sequence of at least two dissociable yet functionally related mental processes. First, observed behaviors must be *identified*. In the context of observing an emotional facial expression (e.g., anger), this necessitates recognizing motor events of the face (e.g., clenching teeth and furrowing the brow). Second, once the expression is understood it can then be *attributed* to an inferred social cause, such as a situational event (e.g., was insulted), mental state (e.g., wants to fight), or disposition (e.g., aggressive person) (Gilbert et al., 1988; Jones and Davis, 1965; Trope, 1986). Given the reliable association of the mirror system with observing, imitating, selectively attending to, and retrieving concrete knowledge about motor actions (Chong et al., 2009; Hesse et al., 2008; Iacoboni, 2009; Spunt et al., 2010), we propose that when faced with an emotional expression, the primary function of the mirror system is the *identification* of emotion-relevant motor events. Conversely, we propose that the mentalizing system, which is reliably associated with representing and reasoning about states of mind (Frith and Frith, 2006; Mitchell, 2009; Saxe, 2006), serves to enable the *attribution* of expressions of emotion to social causes. Critically, such mentalizing-mediated causal attributions are dependent on the prior identification of motor behaviors by the mirror system. Thus, this Identification–Attribution (I–A) model is capable of dissociating the functions of the two systems while at the same time proposing that when causal attributions are made for observed motor behaviors the two systems are functionally related.

Although there is some evidence for the validity I–A model in the context of understanding goal-directed actions (Brass et al., 2007; de Lange et al., 2008; Spunt et al., 2010, 2011), no study to date has explicitly tested an integrative model of mirror and mentalizing system contributions to emotion understanding. In the present study, we used an ecologically valid paradigm (Fig. 1B) for eliciting the explicit identification and attribution of observed emotional expressions. Participants underwent fMRI while viewing short video clips of contextualized emotional responses taken from a dramatic television show. For each clip, we manipulated participants' comprehension goal by having them either identify how the character is showing their feelings (*explicit identification*) or why the character feels as they do (*explicit attribution*). This goal-manipulation allowed use of the same stimuli across conditions.

We tested five hypotheses derived from the I–A model (Fig. 1A). For a given emotional expression (e.g., anger), our first hypothesis was that the mirror system would be preferentially engaged by explicitly identifying emotion-relevant actions of the head and face (Path 2). Our second hypothesis was that the mentalizing system would be preferentially engaged by explicitly attributing observed expressions to a social cause, such as the actor's state of mind in response to something another person has said or done (Path 3).

Whereas the first two hypotheses regard the functional dissociability of the two systems, the remaining hypotheses regard the proposition that the two systems are functionally linked during attributional processing. As illustrated in the model, causal attributions for emotions (Path 2) implicitly demand the prior identification of emotion. Given that this is not an explicit goal of either the identification or attribution tasks, we refer to this as *implicit identification* (Path 1). As argued above, emotion identification depends on encoding motor behaviors of the head and face, a function that previous research suggests is at least partially based in the perception-action matching properties of the mirror system. Hence, our third hypothesis was that areas of the mirror system associated with explicit identification would be active even when participants' explicit goal is to

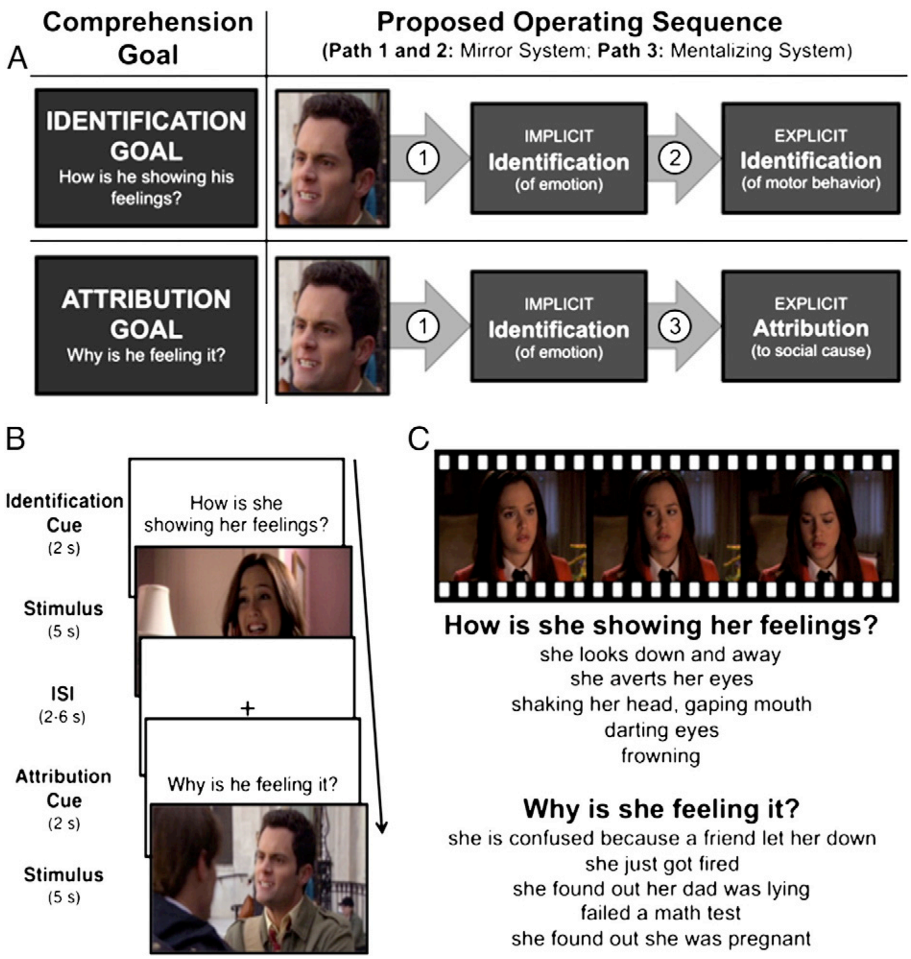


Fig. 1. Identification–Attribution Model and experimental paradigm. (A) Path diagrams illustrating the Identification–Attribution Model of emotion understanding. For the operating sequences, Paths 1 and 2 are hypothesized to rely on the mirror system, while Path 3 is hypothesized to rely on the mentalizing system. (B) Structure of the event-related Why–How task for emotions. (C) Three frames from one stimulus and actual responses for 5 participants to identification and attribution trials.

make attributions, and our fourth hypothesis was that these regions of the mirror system should exhibit activity that is functionally associated with activity in the mentalizing system. Finally, our fifth hypothesis was that, during trials requiring causal attributions, activity in the mirror system should precede activity in the mentalizing system, consistent with existing dual-process models of attribution (Gilbert et al., 1988).

Materials and methods

Participants

Twenty-two participants (12 females, mean age = 21.59, range = 19–32) were recruited from the University of California, Los Angeles (UCLA) subject pool and provided written informed consent according to the procedures of the UCLA Institutional Review Board. All participants were right-handed, native English speakers, metal-free, not claustrophobic, and not taking any psychoactive medications.

Stimuli

Stimuli were taken from season one of the television show *Gossip Girl* (GG). GG was used as a source of stimuli for two reasons: (a) as a professionally filmed dramatic television series, characters in the show frequently exhibit naturalistic emotional responses and (b) the target market for the show includes college-aged individuals, thus increasing the relevance of the stimuli to our college-aged sample. All participants in the present study had not previously seen an episode of GG.

Stimulus selection proceeded as follows. First, candidate clips were cut from episodes of season one of GG. Sound was removed from all clips, and the duration of each clip ranged from 2.5 to 5 s. All clips featured a single camera shot in which a character exhibits an emotional facial expression. In some clips a second character is visible; however, in all cases the main character is the only one facing the camera, and thus the only character whose expression is visible to the viewer. This initial selection phase yielded a sample of over 100 clips. To assess the quality of the clips, we conducted a pilot study in which 26 UCLA undergraduates viewed the clips while seated at a computer station. For each clip, participants rated the valence (forced-choice positive or negative) and intensity of the main character's emotional facial expression. Moreover, participants typed out responses to the following two questions: (a) *How is s/he showing his/her feelings?* and (b) *Why is s/he feeling it?* Response time to these questions was recorded; in addition, participants rated the difficulty they experienced producing each response. All ratings were made using a 1 to 9 Likert scale (anchors: 1 = not at all, 3 = slightly, 5 = somewhat, 7 = very, 9 = extremely). This data was then used to select 40 clips that (a) were reliably categorized as either positive or negative (the final set included 16 positive and 24 negative clips), (b) collectively minimized response time and self-reported difficulty when answering the *how* and *why* questions, and (c) featured a range of characters (15 different characters; 9 males, 31 females). For the final set, the mean intensity rating was 6.13 (SD = 1.42).

Experimental design and procedure

During scanning, participants viewed each of the 40 stimuli twice. For each viewing, they were given either an identification or attribution goal, and the order of goals for each clip was counterbalanced across participants. The structure and timing of the event-related design is displayed in Fig. 1B. Goals were induced prior to stimulus onset by instructing the participant to answer either the question "How is s/he showing his/her feelings?" or the question "Why is s/he feeling it?" (pronouns were adjusted to correspond to the sex of the character in each clip). For *How* trials, participants were asked to describe one part of the person's facial expression or head movement that shows how the they are feeling. For *Why* trials, participants were asked to describe a reason the character might have that would plausibly explain why they feel as they do. For all trials, participants silently thought of their response and made a right index finger button press once they had their response in mind. Prior to the scan, participants were thoroughly trained on a set of 10 clips not featured in the primary task, during which time they performed the task out loud while the experimenter watched. For this training, the following points were emphasized: (a) for all trials, participants were asked not to merely identify the emotion, (b) for *Why* trials, participants were told that there are no right or wrong answers, but that their answer should be plausible given the target's expression, and (c) for *How* trials, participants were told that their responses should be restricted to face and head actions which are clearly part of the emotional response. Following the scan, participants performed the task a second time and typed their responses on a keyboard.

Each trial was separated by a period of variable duration (range = 2–6 s, mean = 3 s) that featured a black screen with a centered fixation cross. The order of trials was optimized for the comparison of *How* and *Why* trials using the OptimizeDesign genetic algorithm (Wager and Nichols, 2003) implemented in MATLAB (The MathWorks, Inc., Natick, MA, USA). The MATLAB Psychophysics Toolbox (Brainard, 1997) was used to present the stimuli to participants and to record their responses. Participants viewed the task through LCD goggles and responded using a four-button box.

Image acquisition

Imaging data were acquired using a Siemens Trio 3.0 Tesla MRI scanner at the UCLA Ahmanson–Lovelace Brainmapping Center. For each participant, we acquired 484 functional T2*-weighted echoplanar image volumes (EPIs; slice thickness = 3 mm, gap = 1 mm, 36 slices, TR = 2000 ms, TE = 25 ms, flip angle = 90°, matrix = 64 × 64, FOV = 200 mm) divided evenly across two runs. We also acquired a T2-weighted matched-bandwidth anatomical scan (same parameters as EPIs, except: TR = 5000 ms, TE = 34 ms, flip angle = 90°, matrix = 128 × 128) and a T1-weighted magnetization-prepared rapid-acquisition gradient echo anatomical scan (slice thickness = 1 mm, 176 slices, TR = 2530 ms, TE = 3.31 ms, flip angle = 7°, matrix = 256 × 256, FOV = 256 mm).

Behavior analysis

MATLAB was used to analyze all behavioral data. For each participant, we computed the mean response time for each condition, and used a paired-sample *t*-test to assess the significance of the difference. Due to technical difficulties, response time data was not available for one participant. In addition, we combined all participants' post-scan responses and computed the frequency of each word in the *How* and *Why* conditions separately. Words were then sorted by frequency in order to determine the concepts most commonly used as a function of comprehension goal.

Image analysis

Functional data were analyzed using Statistical Parametric Mapping (SPM5, Wellcome Department of Cognitive Neurology, London, UK) operating in MATLAB. Within each run, image volumes were realigned to correct for head motion; normalized into Montreal Neurological Institute space (re-sampled at 3 × 3 × 3 mm); and smoothed with an 8 mm Gaussian kernel, full width at half maximum.

To model the effects of comprehension goal on the BOLD response to the stimuli, we setup a general linear model for each participant. Regressors for *How* and *Why* trials were created by convolving a delta function at stimulus onset with a canonical (double-gamma) hemodynamic response function (HRF). In addition, each model included several covariates of no interest. For each trial type, we included parametric modulators of the height of the predicted BOLD response as a function of three trial-variable parameters: (a) response time, (b) stimulus valence, and (c) stimulus intensity. The latter two parameters were produced from ratings collected in the pilot study described above. Each parametric regressor was created by multiplying the delta functions at each onset by the de-meaned parameter values, and then convolving the resulting vector with the canonical HRF. Additional covariates included regressors modeling participants' button-presses, skipped trials, and the six motion parameters. For each model, the timeseries was high-pass filtered to 1/100 Hz, and serial autocorrelations were modeled as an AR(1) process.

Following estimation, contrast images were created for the following comparisons: *How* > *Fixation*, *Why* > *Fixation*, *How* > *Why*, and *Why* > *How*. All contrast images were then entered into a second-level analysis. To assess areas of the brain sensitive to the observer's goal, we used one-sample *t*-tests to determine areas present in the contrasts *How* > *Why* and *Why* > *How*. To assess areas of the brain engaged by the attribution goal that were preferentially engaged by the identification goal, we first created a mask of regions more active in the comparison *How* > *Why*, corrected using a false discovery rate (FDR) of .05 combined with an extent threshold of 20 voxels. This mask was then used to interrogate *Why* > *Fixation*.

We used psychophysiological interactions (PPIs; Friston et al., 1997) to assess whether areas of the mirror system showed functional connectivity with areas of the mentalizing system during performance of the task. PPI enables determination of brain regions that show a change in correlation with a seed region (the "physiological" component of the PPI) as a function of a change in participants' psychological state (the "psychological" component of the PPI). As seeds, we used the 5 clusters identified in the analysis of *Why* > *Fixation* masked by *How* > *Why* (Fig. 3). We then setup five PPI models for each participant, one for each seed region. Each model included two PPIs, one for the effect of *How* (versus fixation) and one for the effect of *Why* (versus fixation). PPI regressors were created in the following way: (a) for each participant, we first defined the timeseries of the seed region as the first eigenvariate (adjusting for the motion regressors and session means); (b) the timeseries was deconvolved to estimate the underlying neural activity using the deconvolution algorithm in SPM5 (Gitelman et al., 2003); (c) the deconvolved timeseries was multiplied by the predicted time series (pre-convolved) of each condition, resulting in one "neural" PPI for each condition, and (d) each neural PPI was then convolved with the canonical HRF, yielding the two PPI regressors. As covariates of no interest, each model also included the timeseries of the seed region, the convolved timeseries of each condition, and the six motion parameters. The timeseries was high-pass filtered to 1/100 Hz, and serial autocorrelations were modeled as an AR(1) process.

After estimation, contrast images of the PPI effects (*Why* > *Fixation*, *How* > *Fixation*, *Why* > *How*) were then subjected to second-level one-sample *t*-tests to enable group inference. Given our *a priori* interest in determining connectivity of our seed regions with areas associated with attributional processing, we used the contrast of *Why*>*How* (correct using a FDR of .001 combined with an extent threshold of 10 voxels) to inclusively mask these group analyses. Thus, voxels showing a positive PPI effect can be interpreted as areas preferentially engaged by the attribution goal that exhibit increased effective connectivity with areas preferentially engaged by the identification goal.

All main effect analyses were thresholded using an FDR of .001 combined with an extent threshold of 10 voxels. To increase sensitivity for detecting interaction effects, all PPI analyses used a less conservative FDR correction of .05 combined with an extent threshold of 20 voxels. For visual presentation, thresholded *t*-statistic maps were either (a) surface rendered using the SPM5 Surfrend toolbox Version 1.0.2 (I. Kahn; http://spmsurfrend.sourceforge.net) and overlaid on a surface-based representation of the MNI canonical brain using the NeuroLens analysis package (Hoge and Lissot, 2004), or (b) overlaid on the average of the participants' T1-weighted anatomical images. For graphing purposes, percent signal change was calculated using the rfxplot toolbox (Gläscher, 2009).

Rfxplot was used to compute peri-stimulus time histograms (PSTHs) of the mean event-related response to *Why* trials in the right pIFG ROI and each region observed in the PPI analysis displayed in Fig. 4A. The PSTHs spanned the peri-stimulus period -2 to 10 s, and data was split in 2 s time bins corresponding to the TR. To investigate whether the peak response occurred significantly later in mentalizing than in mirror regions, we defined the time of peak in each region on a subject-by-subject basis as the bin containing the maximum value in the peri-stimulus period 2 to 10 s. Paired sample t-tests were then used to compare the time of peak in the right pIFG ROI to each of the mentalizing ROIs.

Results

Behavioral results

A paired-samples *t*-test showed that participants took longer to respond to *Why* trials (M = 3.84 s, SD = 1.01) than to *How* trials (M = 3.49 s, SD = .90), $t_{20}=3.338$, $p=.003$. In our previous studies using variants of the Why–How task (Spunt et al., 2010, 2011), response time was shown to have little to no effect on the differential response to *How* and *Why* trials. To be conservative we have included response time as a covariate of no interest in the analyses presented below.

We determined the effect of the goal manipulation (*How* vs. *Why*) by examining the words most frequently used in the two conditions. Sample responses to one stimulus are featured in Fig. 1C. For *How* trials, participants most frequently use nouns referencing parts of the head and face ("eye", "mouth", "head"), verbs indicating actions of the face ("look", "smile", "open", "cry"), and adverbs indicating movement through space ("down", "up"). For *Why* trials, participants most frequently use words indicating that they attributed the emotional response to a recent event (adverb "just" and noun "news"), which typically was something another character had said or done (nouns "friend", "boyfriend" and "girlfriend"; preposition "with"; verb "told"). Finally, the conjunction "because" was frequently used during *Why* trials, confirming the presence of causal attribution. These data face validly confirm the manipulation, namely, that during *How* trials, participants attended to and named emotion-relevant actions of the head and face, and during *Why* trials, participants inferred a plausible cause of the observed emotional response.

Neuroimaging results

Our first hypothesis was that the mirror system would be preferentially engaged by the explicit identification goal. We tested this with the contrast *How*>*Why*. As displayed in Fig. 2 and listed in Table 1, this revealed robust activation in the core regions of the mirror system for actions: bilateral pIFG, left dPMC, and bilateral rIPL. We also observed bilateral activation in areas within the lateral occipitotemporal cortex and the superior parietal lobule. These results demonstrate that the mirror, and not the mentalizing system, subserves the explicit identification of emotion-relevant motor behaviors.

Our second hypothesis was that the mentalizing system would be preferentially engaged by the explicit attribution goal. We tested this with the contrast *Why*>*How*. As displayed in the lower part of Fig. 2 and listed in Table 1, this revealed robust activity in areas of the mentalizing system, including dmPFC, vmPFC, and PCC/PC on the cortical midline, and bilateral activations in TPJ, pSTS and aTC. Additional areas of activation included bilateral ventrolateral prefrontal cortex, pre-supplementary motor area, mid superior temporal sulcus, and bilateral parahippocampal cortex. These results demonstrate that the mentalizing system, and not the mirror system, subserves the explicit attribution of an observed emotional expression to a social cause.

These analyses indicate that in the context of emotion understanding, explicit identification and attribution goals clearly dissociate activation in the mirror and mentalizing systems, respectively. However, the I–A Model also specifies a functional linkage between the two systems during attributional processing, such that the mirror system contributes to the rapid identification of expressions that can then serve as inputs to attributional processing in the mentalizing system. Hence, our third hypothesis was that areas of the mirror system associated with identifying motor behavior would be active even when participants' explicit goal was to make causal attributions. To test this, we interrogated the contrast *Why*>*Fixation* masked by only those voxels preferentially engaged by the explicit identification goal (i.e., *How*>*Why*). This analysis reveals identification-related regions that display robust above-baseline activity even during explicit

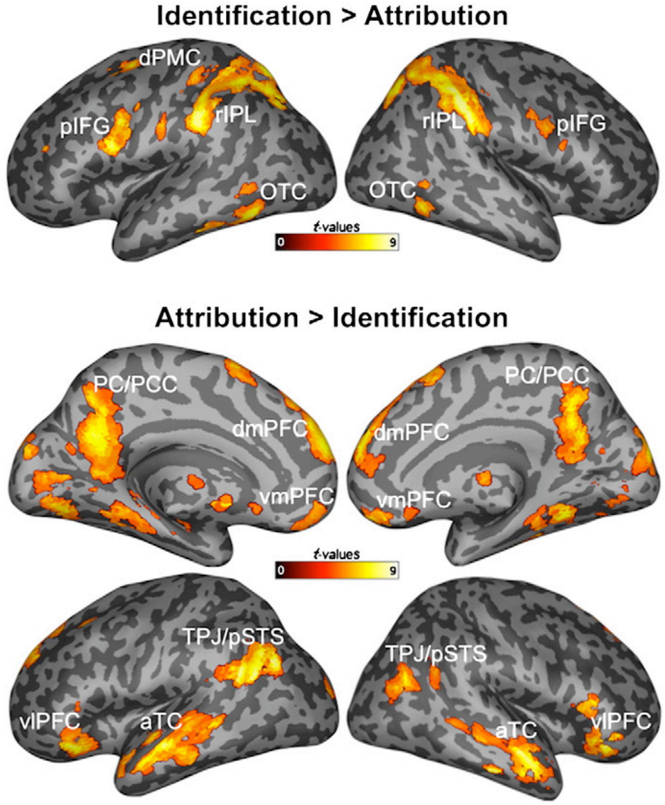


Fig. 2. Regions associated with explicit identification and attribution (FDR corrected at .001 across the whole-brain with an extent threshold of 10 voxels). The top row depicts the comparison *How*>*Why*, while the bottom two rows depicts the comparison *Why*>*How*.

Table 1
All regions observed in the main effects of observer goal.

		MNI				
Anatomical region	L/R	x	y	z	t	k
Identify how > infer why						
Frontal cortex						
Ventral premotor cortex/IFG (opercularis)	L	−51	6	27	8.92	111
	R	54	6	30	6.01	53
Dorsal premotor cortex	L	−27	−6	48	6.65	63
Dorsolateral prefrontal cortex	L	−48	42	18	5.91	10
Mid cingulate cortex	L/R	0	−3	30	6.60	20
Parietal cortex						
Rostral inferior parietal lobule	R	54	−33	45	10.77	744_a
	L	−63	−27	36	8.92	680_b
Superior parietal lobule	R	27	−69	51	8.14	744_a
	L	−24	−66	39	7.60	680_b
Anterior intraparietal sulcus	L	−30	−48	54	7.63	680_b
Temporal cortex						
Posterior inferior/middle temporal gyrus	L	−54	−63	−12	7.87	94_c
	L	−51	−39	−18	6.43	94_c
	R	57	−60	−6	7.12	37
Cerebellum						
Cerebellum (posterior lobe)	R	18	−72	−45	5.78	14
Infer why > identify how						
Frontal cortex						
Inferior frontal gyrus (orbitalis)	L	−42	27	−12	8.09	129
	R	45	27	−12	7.99	127_a
Inferior frontal gyrus (triangularis)	R	57	27	9	6.51	127_a
Dorsomedial prefrontal cortex (BA 8/9)	L	−12	51	33	7.84	536_b
	R	3	63	24	7.52	536_b
	R	6	45	51	6.33	536_b
Medial prefrontal cortex (BA 10)	R	12	51	9	5.11	536_b
Ventromedial prefrontal cortex BA 11)	R	3	57	−18	6.76	109
Pre-supplementary motor area	R	6	21	60	6.65	98
Dorsolateral prefrontal cortex (BA 8)	L	−27	24	45	5.77	12
Subgenual anterior cingulate cortex	R	6	18	−18	5.48	12
Parietal cortex						
Temporoparietal junction	L	−48	−54	24	9.39	250_c
	L	−45	−69	39	9.05	250_c
	R	54	−66	24	8.55	140
Precuneus/posterior cingulate cortex	L	−3	−60	27	7.98	673_d
	R	12	−54	39	7.12	673_d
Retrosplenial cortex	L	−18	−54	9	6.39	673_d
Temporal cortex						
Mid superior temporal sulcus	R	51	−18	−21	9.82	242_e
	L	−57	−12	−9	8.13	340_f
Posterior superior temporal sulcus	R	−57	−39	3	5.63	242_e
	L	−63	−39	3	6.52	340_f
Anterior superior temporal sulcus	L	−54	12	−21	6.44	340_f
Parahippocampal cortex	R	30	−48	−6	7.42	82
	L	−21	−39	−9	7.21	62
Occipital cortex						
Cuneus	R	12	−96	21	7.34	356_c
	L	−9	−93	21	6.82	356_c
Lingual gyrus	L	−21	−75	−9	6.56	356_c
	R	15	−75	−6	5.09	18
Cerebellum						
Cerebellum (posterior lobe)	L/R	0	−57	−45	5.80	20
	L	−30	−78	−33	5.49	41
	R	30	−78	−33	5.40	19
Subcortical						
Ventral striatum	L	−9	9	−9	6.69	15
Thalamus	R	3	−6	9	5.86	24
Hippocampus	L	−27	−12	−18	5.38	18

Note. N = 22. All regions FDR corrected at .001. Coordinates are all local maxima observed which were separated by at least 20 mm. L/R = left and right hemispheres; x, y, and z = Montreal Neurological Institute (MNI) coordinates in the left–right, anterior–posterior, and inferior–superior dimensions, respectively; *t* = *t* statistic value at those coordinates; *k* = cluster voxel extent (coordinates with *k*s that share the same subscript originate from the same cluster); BA = Brodmann's Area; IFG = Inferior frontal gyrus.

attributional processing. As displayed in Fig. 3 and listed in Table 2, this analysis revealed bilateral activations in canonical areas of the frontal mirror system, namely posterior inferior frontal gyrus bordering the precentral sulcus, as well as activation in bilateral posterior inferior temporal gyrus and left posterior middle temporal gyrus. As is apparent from the graphs in Fig. 3, these regions are preferentially engaged by explicit identification *but are also* strongly engaged during attributional processing. Given this, we suggest these regions contribute to both the identification of motor behavior explicitly demanded by *How* trials, and the identification of emotional expression implicitly demanded by both *How* and *Why* trials.

Our fourth hypothesis sought to directly test for a functional association between the mirror and mentalizing systems during emotion comprehension. We used psychophysiological interaction (PPI) analyses to test this hypothesis, using the mirror and visual areas displayed in Fig. 3 as seed regions. For all seeds, we found no evidence of effective connectivity with mentalizing regions during *How* trials. However, during *Why* trials, we did find evidence of effective connectivity between only the right pIFG seed and core areas of the mentalizing system: dmPFC, vmPFC, PCC/PC, bilateral TPJ, and left aTC (Fig. 4A; coordinates listed Table 3). No regions emerged when statistically comparing effective connectivity with right pIFG (or any of the other seeds) across *Why* and *How* trials, a null result which we consider in the Discussion.

In addition to predicting a functional association, the I–A model proposes a specific sequence of operations during attributional inference, with behavior identification occurring first followed by causal attribution (Fig. 1A). Thus, our final hypothesis was that mirror system activity should precede activity in the mentalizing system. To garner evidence for this operating sequence, we compared the time-course of the event-related hemodynamic response to *Why* trials in right pIFG to each of the mentalizing regions observed in the PPI analysis. Of the five regions demonstrating a functional association with right pIFG, all but ventromedial prefrontal cortex exhibit a significantly later time to peak, all $ts(22) > 2.81$, $ps < .02$ (Fig. 4B). Along with the observation of effective connectivity among right pIFG and these regions, we take this to suggest that mirror system activity both precedes *and* informs activity in the mentalizing system.

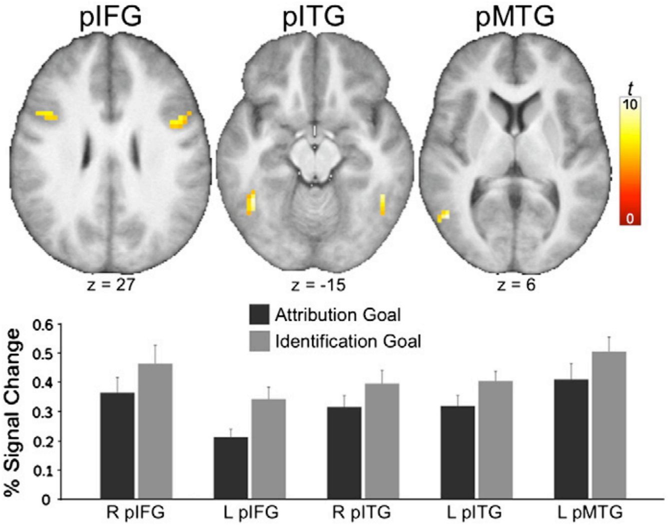


Fig. 3. Identification-related regions that exhibit above-baseline activity in response to the attribution goal. Statistical parametric maps are from the contrast *Why* > *Fixation* masked with *How* > *Why* (FDR corrected at .001 with an extent threshold of 10 voxels). Graphs display percent signal change from fixation baseline in the five regions as a function of comprehension goal.

Table 2
All regions observed in why > fixation masked by how > why.

Anatomical region	L/R	MNI x	MNI y	MNI z	t	k
Frontal cortex						
Posterior inferior frontal gyrus	L	−51	12	21	8.22	25
	R	48	9	27	7.20	22
Occipitotemporal cortex						
Posterior middle temporal gyrus	L	−51	−63	6	9.49	16
Posterior inferior temporal gyrus	L	−45	−51	−15	9.19	26
	R	48	−48	−15	8.28	42

Note. N = 22. All voxels FDR corrected at .001. Coordinates are all local maxima observed which were separated by at least 20 mm. L/R = left and right hemispheres; x, y, and z = Montreal Neurological Institute (MNI) coordinates in the left–right, anterior–posterior, and inferior–superior dimensions, respectively; *t* = *t* statistic value at those coordinates; *k* = cluster voxel extent (coordinates with *k*s that share the same subscript originate from the same cluster).

Discussion

We tested an integrative I–A model of the brain systems supporting emotion understanding. Using a method for inducing naturalistic emotion comprehension, we found evidence for dissociable yet functionally related roles for the mirror and mentalizing systems. We observed that the mirror, but not the mentalizing system, was recruited by behavior identification whether such identification was an explicit goal or merely an implicit requirement of the task. In contrast, we observed that the mentalizing, but not the mirror system, was recruited by causal attributions for observed emotional expressions. In addition to dissociating the function of the two systems, we report evidence that during attributional processing, mirror system activity both precedes and is functionally associated with activity in the mentalizing system. Taken together, these results provide strong empirical support for an integrative model of mirror and mentalizing system contributions to emotion understanding.

A fundamental strength of the present study that separates it from past research on emotion understanding is the ecological validity of the methods used. Past research on the neural bases of emotion understanding has typically relied on stimuli and tasks that sacrifice ecological validity for the sake of experimental control (for a discussion of this, see Zaki and Ochsner, 2009). For stimuli, we used clips from a professionally produced television show that features actors who are expert at producing the types of emotional expressions that individuals encounter in the social world outside of the scanner. Moreover, participants are aware that scenes are extracted from a rich ongoing narrative drama, making attributional processing natural. Though fictional, narrative media such as television, cinema, and literature produce in the observer a simulation of actual social experience (Mar and Oatley, 2008) often resulting in the induction of strong social emotions in the observer (Gardner and Knowles, 2008). The participants' task was also ecologically valid, using natural language to induce spontaneous, open-ended social cognition. The interrogatives *how* and *why*, which frequently appear in everyday discourse, are face valid inductions of identification and attribution goals. Superficially, it

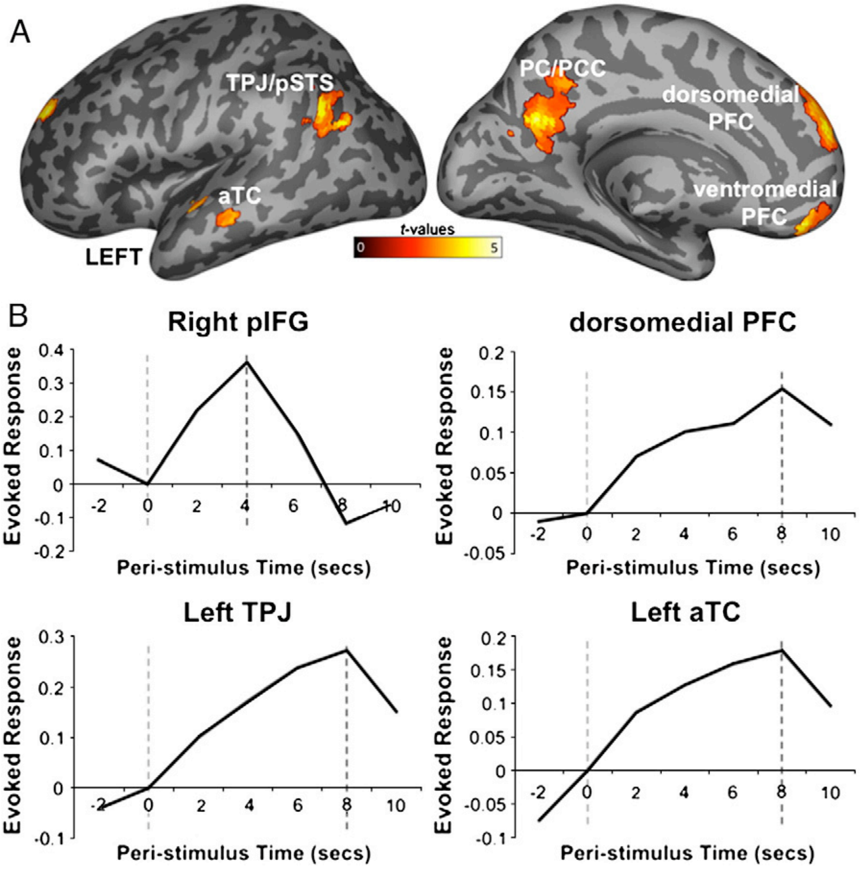


Fig. 4. Functional association of mirror and mentalizing systems during attributional processing. (A) Mentalizing regions showing positive effective connectivity, using psychophysiological interactions (PPI), with right pIFG in the comparison *Why > Fixation* (this analysis was inclusively masked with main effect of explicit attribution, and is FDR corrected at .05 with an extent threshold of 20 voxels). (B) Peri-stimulus timecourse histograms of right pIFG and the regions of the mentalizing system observed in (A). The light dashed vertical line marks time of stimulus onset, and the darker dashed line marks average time-of-peak.

Table 3
All regions observed in the psychophysiological interaction analysis.

		MNI				
Anatomical region	L/R	x	y	z	t	k
Right pIFG seed						
Frontal cortex						
Dorsomedial Prefrontal cortex (BA 8/9)	R	9	48	39	4.77	258_a
	L/R	0	63	24	4.48	258_a
Medial prefrontal cortex (BA 10)	R	12	48	12	3.27	258_a
Ventromedial prefrontal cortex	L/R	0	45	−15	4.71	77
Temporal cortex						
Anterior temporal cortex	L	−57	−6	−6	3.95	33
Parietal cortex						
Precuneus/posterior Cingulate cortex	L	−6	−63	24	4.42	152_b
	L	−12	−48	39	3.10	152_b
Temporoparietal junction	L	−51	−66	33	4.99	77
	R	54	−66	33	3.37	31
Left pIFG seed						
No suprathreshold voxels						

Note. $N=22$. All voxels FDR corrected at .05. Coordinates are all local maxima observed which were separated by at least 20 mm. L/R = left and right hemispheres; x, y, and z = Montreal Neurological Institute (MNI) coordinates in the left–right, anterior–posterior, and inferior–superior dimensions, respectively; $t=t$ statistic value at those coordinates; k = cluster voxel extent (coordinates with ks that share the same subscript originate from the same cluster); BA = Brodmann's Area; vPMC = Ventral premotor cortex.

may appear that an open-ended response format combined with complex stimuli would produce unreliable BOLD data. However, the effects reported in the present study are very reliable, both within and across participants.

In fact, we speculate that the ecological validity of both our task and stimuli explains why this is the first study to observe a functional association between mirror and mentalizing system activity during emotion understanding. Extant research on the mirror system has almost exclusively relied on tasks involving the passive observation or active imitation of concrete motor actions. In the context of the I–A model (Fig. 1A), such methods isolate processes involved in the comprehension of motor behaviors (Paths 1 and 2), but provide little opportunity for isolating processes involved in causal attribution (Path 3). The opposite can be said of research on the mentalizing system, which typically has manipulated the extent to which participants represent states of mind (Path 3), but has done so with ecologically limited stimuli that do not depict real motor behaviors (hence avoiding Path 1), and instead feature verbal, abstract, or contextually impoverished depictions of human behaviors. In the present study, we explicitly induced causal attributions for realistic and highly contextualized emotional stimuli, and found evidence that mirror and mentalizing systems cooperate in this condition. Conversely, we found no evidence of effective connectivity amongst the two systems during performance of the explicit identification task.

As formulated in Fig. 1, the I–A Model does not predict effective connectivity amongst the mirror and mentalizing systems during performance of the explicit identification task. Hence, it might be predicted that connectivity amongst the systems should be stronger during *Why* trials than during *How* trials. However, direct comparison of *Why* to *How* trials revealed no significant differences in effective connectivity amongst either of the pIFG seeds and the mentalizing system. We make two points regarding this null result. First, at exploratory statistical thresholds ($p<.05$ uncorrected, voxel extent = 20), we do observe preliminary evidence of increased mirror-mentalizing coupling during *Why* compared to *How* trials. Second, it is already known that mentalizing system activity can occur spontaneously during the perception of social stimuli (Brass et al., 2007; Castelli et al., 2000; Iacoboni et al., 2004; Ma et al., 2011); in parallel, it is well documented in social psychology that attributional processing can occur spontaneously (Uleman et al., 2007). Hence, it is plausible that some degree of spontaneous attributional processing may accompany the identification goal during *How* trials, and such spontaneous attributional processing may rely on spontaneous coupling of the two systems' function during social perception. Further investigation is required to determine whether such spontaneous coupling exists, as well as whether the degree of functional coupling is modulated by the observer's comprehension goal.

We observed that identification and attribution goals strongly distinguished activity in the mirror and mentalizing systems, respectively. Such a strong dissociation is consistent with past research on the neural bases of emotion understanding (Shamay-Tsoory, 2011). One influential model distinguishes two types of emotion understanding: (a) emotional empathy, which involves sharing the affective states of others and is dependent on perception–action matching in the mirror system (and downstream effects in brain areas for the production of affective experience), and (b) cognitive empathy, which involves adopting the perspective of others and is dependent on the representation of mental states in the mentalizing system. It has been shown that whereas damage to frontal mirror areas is selectively associated with reports of trait emotional empathy, damage to mentalizing system areas (namely, vmPFC or pSTS) is selectively associated with a lack of trait cognitive empathy (Shamay-Tsoory et al., 2009). Thus, there are clear conceptual parallels between the I–A model and the distinction between emotional and cognitive empathy, with the former overlapping with identification and the latter with causal attribution (see Waytz and Mitchell, 2011 for similar theoretical distinction). However, the I–A model is distinguishable in at least two ways. First, the I–A model not only dissociates the mirror and mentalizing systems, but also specifies the nature of their functional relationship. Second, the I–A model is not tied to a particular category of social stimulus, but can and has been applied to other categories of social stimuli, such as goal-directed motor actions (Spunt et al., 2010, 2011).

Both the mirror and mentalizing systems have been implicated in psychopathologies that feature impairments in emotion understanding, particularly autism spectrum disorder (ASD). One prominent hypothesis is that these are caused by a dysfunctional mirror neuron system (Oberman and Ramachandran, 2007). Although this hypothesis has received some empirical support, the extant data is far from conclusive (Baird et al., 2011; Decety, 2010; Southgate and Hamilton, 2008), and indeed even the evidence for emotion recognition deficits in autism is mixed. In fact, the extant research suggests there is no reliable neural correlate of the social deficits in ASD (Amaral et al., 2008). The results of the present study, and the I–A model more generally, provide novel perspective on the relationship between brain function and the neural bases of individual differences in social cognitive functioning. More specifically, the I–A model distinguishes three independent determinants, each of which are sufficient for explaining variance in social cognitive outcomes. One determinant may indeed be mirror system functioning; in this case, dysfunction would be expected to relate to performance on tasks requiring the identification of motor behaviors either explicitly (e.g., naming or imitating facial expressions) or implicitly (e.g., causal attribution for facial expressions of emotion). However, causal attribution for social stimuli that do not require the identification of motor behaviors (e.g., verbal description of an emotion) would be predicted to be relatively unaffected. A second determinant may be mentalizing system function; in this case, dysfunction would expected to impair performance on social cognitive tasks requiring causal attribution, regardless of the nature of the social stimulus, but would leave performance on tasks requiring the identification of motor behaviors relatively unaffected. Finally, a third determinant may be the functional integration of mirror and mentalizing system activity; in this case, dysfunction would be expected to impair performance only on tasks that require causal attributions for sensory depictions of motor behavior. In this way, the I–A model suggests fruitful directions for future research on the complex and nuanced neural basis of both normal and abnormal social cognition.

Acknowledgments

The authors thank Edward Kobayashi and John Mezzanotte for their assistance performing the study, Marco Iacoboni and Sylvia Morelli for helpful advice, and several anonymous reviewers for helpful feedback on an earlier version of the manuscript. For generous support the authors thank the Brain Mapping Medical Research Organization, the Brain Mapping Support Foundation, the Pierson-Lovelace Foundation, The Ahmanson Foundation, William M. and Linda R. Dietel Philanthropic Fund at the Northern Piedmont Community Foundation, the Tamkin Foundation, the Jennifer Jones-Simon Foundation, the Capital Group Companies Charitable Foundation, Robson Family and the Northstar Fund.

References

Amaral, D.G., Schumann, C.M., Nordahl, C.W., 2008. Neuroanatomy of autism. Trends Neurosci. 31, 137–145.

Baird, A., Scheffer, I., Wilson, S., 2011. Mirror neuron system involvement in empathy: a critical look at the evidence. Soc. Neurosci. 2011, 1–9.

Bastiaansen, J., Thioux, M., Keysers, C., 2009. Evidence for mirror systems in emotions. Philos. Trans. R. Soc. Lond. B 364, 2391.

Blair, R.J.R., 2005. Responding to the emotions of others: dissociating forms of empathy through the study of typical and psychiatric populations. Conscious. Cogn. 14, 698–718.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 443–446.

Brass, M., Schmitt, R.M., Spengler, S., Gergely, G., 2007. Investigating action understanding: inferential processes versus action simulation. Curr. Biol. 17 (24), 2117–2121.

Buccino, G., Binkofski, F., Fink, G.R., et al., 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur. J. Neurosci. 13, 400–404.

Budell, L., Jackson, P., Rainville, P., 2010. Brain responses to facial expressions of pain: emotional or motor mirroring? Neuroimage 53, 355–363.

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J.C., Lenzi, G.L., 2003. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. Proc. Natl. Acad. Sci. 100, 5497–5502.

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

Chong, T.T.-J., Cunnington, R., Williams, M.A., Mattingley, J.B., 2009. The role of selective attention in matching observed and executed actions. Neuropsychologia 47 (3), 786–795.

de Lange, F.P., Spronk, M., Willems, R.M., Toni, I., Bekkering, H., 2008. Complementary systems for understanding action intentions. Curr. Biol. 18 (6), 454–457.

Decety, J., 2010. To what extent is the experience of empathy mediated by shared neural circuits? Emotion Rev. 2, 204–207.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. Exp. Brain Res. 91 (1), 176–180.

Dimberg, U., Thunberg, M., Elmehed, K., 2000. Unconscious facial reactions to emotional facial expressions. Psychol. Sci. 11, 86.

Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: from action organization to intention understanding. Science 308 (5722), 662–667.

Fox, M., Snyder, A., Vincent, J., et al., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. 102, 9673.

Friston, K., Buechel, C., Fink, G., et al., 1997. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6, 218–229.

Frith, C.D., Frith, U., 2006. The neural basis of mentalizing. Neuron 50, 531–534.

Gallese, V., 2007. Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. Philos. Trans. R. Soc. Lond. B 362, 659–669.

Gallese, V., Keysers, C., Rizzolatti, G., 2004. A unifying view of the basis of social cognition. Trends Cogn. Sci. 8, 396–403.

Gardner, W., Knowles, M., 2008. Love makes you real: favorite television characters are perceived as "real" in a social facilitation paradigm. Soc. Cogn. 26 (2), 156–168.

Gilbert, D.T., 1998. In: Gilbert, D.T., Fiske, S.T., Lindzey, G. (Eds.), The Handbook of Social Psychology, 4th ed. McGraw, New York, pp. 89–150.

Gilbert, D., Pelham, B., Krull, D., 1988. On cognitive busyness: when person perceivers meet persons perceived. J. Pers. Soc. Psychol. 54, 733–740.

Gitelman, D., Penny, W., Ashburner, J., Friston, K., 2003. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19 (1), 200–207.

Gläscher, J., 2009. Visualization of group inference data in functional neuroimaging. Neuroinformatics 7, 73–82.

Hesse, M., Sparing, R., Fink, G., 2008. End or means-the "what" and "how" of observed intentional actions. J. Cogn. Neurosci. 21 (4), 776–790.

Heyes, C., 2010. Mesmerising mirror neurons. Neuroimage 51 (2), 789–791.

Hickok, G., 2009. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. J. Cogn. Neurosci. 21 (7), 1229–1243.

Hoge, R.D., Lissot, A., 2004. NeuroLens: an integrated visualization and analysis platform for functional and structural neuroimaging. Proc. Int. Soc. Magn. Reson. Imaging 11, 1096.

Iacoboni, M., 2009. Imitation, empathy, and mirror neurons. Annu. Rev. Psychol. 60, 653–670.

Iacoboni, M., Lieberman, M., Knowlton, B., et al., 2004. Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. Neuroimage 21, 1167–1173.

Jacob, P., 2008. What do mirror neurons contribute to human social cognition? Mind Lang. 23 (2), 190–223.

Jacob, P., Jeannerod, M., 2005. The motor theory of social cognition: a critique. Trends Cogn. Sci. 9, 21–25.

Jones, E.E., Davis, K.E., 1965. From acts to dispositions: the attribution process in person perception. In: Berkowitz, L. (Ed.), Advances in Experimental Social Psychology, vol. 2. Academic Press, New York, pp. 220–266.

Kelley, H., 1973. The processes of causal attribution. Am. Psychol. 28 (2), 107–128.

Keysers, C., Gazzola, V., 2007. Integrating simulation and theory of mind: from self to social cognition. Trends Cogn. Sci. 11, 194–196.

Lombardo, M.V., Chakrabarti, B., Bullmore, E.T., et al., 2010. Shared neural circuits for mentalizing about the self and others. J. Cogn. Neurosci. 22, 1623–1635.

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 (2), 123–138.

Mar, R.A., Oatley, K., 2008. The function of fiction is the abstraction and simulation of social experience. Pers. Psychol. Sci. 3, 173–192.

Mitchell, J., 2009. Inferences about mental states. Philos. Trans. R. Soc. B 364 (1521), 1309.

Mitchell, J., Macrae, C., Banaji, M., 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50, 655–663.

Mukamel, R., Ekstrom, A.D., Kaplan, J., Iacoboni, M., Fried, I., 2010. Single-neuron responses in humans during execution and observation of actions. Curr. Biol. 20 (8), 750–756.

Neal, D.T., Chartrand, T.L., 2011. Embodied emotion perception: amplifying and dampening facial feedback modulates emotion perception accuracy. Soc. Psychol. Pers. Sci. 2011, 1–7.

Niedenthal, P.M., Mermillod, M., Maringer, M., Hess, U., 2010. The Simulation of Smiles (SIMS) model: embodied simulation and the meaning of facial expression. Behav. Brain Sci. 33, 417–433.

Oberman, L.M., Ramachandran, V.S., 2007. The simulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. Psychol. Bull. 133, 310–327.

Ochsner, K.N., Knierim, K., Ludlow, D.H., et al., 2004. Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. J. Cogn. Neurosci. 16, 1746–1772.

Olsson, A., Ochsner, K.N., 2008. The role of social cognition in emotion. Trends Cogn. Sci. 12, 65–71.

Preston, S., De Waal, F., 2001. Empathy: its ultimate and proximate bases. Behav. Brain Sci. 25, 1–20.

Saxe, R., 2005. Against simulation: the argument from error. Trends Cogn. Sci. 9, 174–179.

Saxe, R., 2006. Uniquely human social cognition. Curr. Opin. Neurobiol. 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 9, 1835–1842.

Shamay-Tsoory, S.G., 2011. The neural bases for empathy. Neuroscientist 17, 18–24.

Shamay-Tsoory, S., Aharon-Peretz, J., Perry, D., 2009. Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain 132, 617–627.

Southgate, V., Hamilton, A.F., 2008. Unbroken mirrors: challenging a theory of Autism. Trends Cogn. Sci. 12, 225–229.

Spengler, S., Cramon, D.Y.V., Brass, M., 2009. Control of shared representations relies on key processes involved in mental state attribution. Hum. Brain Mapp. 30, 3704–3718.

Spunt, R.P., Falk, E.B., Lieberman, M.D., 2010. Dissociable neural systems support retrieval of how and why action knowledge. Psychol. Sci. 21, 1593–1596.

Spunt, R.P., Satpute, A.B., Lieberman, M.D., 2011. Identifying the what, why, and how of an observed action: an fMRI study of mentalizing and mechanizing during action observation. J. Cogn. Neurosci. 23, 63–74.

Tkach, D., Reimer, J., Hatsopoulos, N.G., 2007. Congruent activity during action and action observation in motor cortex. J. Neurol. 27 (48), 13241–13250.

Trope, Y., 1986. Identification and inferential processes in dispositional attribution. Psychol. Rev. 93, 239–257.

Uleman, J., Saribay, S., Gonzalez, C., 2007. Spontaneous inferences, implicit impressions, and implicit theories. Annu. Rev. Psychol. 59, 329–360.

Van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage 48, 564–584.

Wager, T.D., Nichols, T.E., 2003. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. Neuroimage 18, 293–309.

Waytz, A., Mitchell, J.P., 2011. Two mechanisms for simulating other minds: dissociations between mirroring and self-projection. Curr. Dir. Psychol. Sci. 20, 197–200.

Zaki, J., Ochsner, K., 2009. The need for a cognitive neuroscience of naturalistic social cognition. Ann. N. Y. Acad. Sci. 1167, 16–30.

Zaki, J., Ochsner, K., 2011. Reintegrating the study of accuracy into social cognition research. Psychol. Inquiry 22, 159–182.

Zaki, J., Weber, J., Bolger, N., Ochsner, K., 2009. The neural bases of empathic accuracy. Proc. Natl. Acad. Sci. 106, 11382–11387.

Zaki, J., Hennigan, K., Weber, J., Ochsner, K.N., 2010. Social cognitive conflict resolution: contributions of domain-general and domain-specific neural systems. J. Neurosci. 30, 8481–8488.