Beyond the Language Given: The Neural Correlates of Inferring Speaker Meaning

Jana Bašnáková[1,2], Kirsten Weber[3], Karl Magnus Petersson[1], Jos van Berkum[1,4] and Peter Hagoort[1,3]

[1]Max Planck Institute for Psycholinguistics, Nijmegen 6525 XD, The Netherlands, [2]Institute of Experimental Psychology, SAS, Bratislava 813 64, Slovakia, [3]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen 6525 EN, The Netherlands and [4]Utrecht Institute for Linguistics OTS, Trans 10, 3512 JK, Utrecht, The Netherlands

Address correspondence to Peter Hagoort, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands. Email: p.hagoort@mpi.nl

Even though language allows us to say exactly what we mean, we often use language to say things indirectly, in a way that depends on the specific communicative context. For example, we can use an apparently straightforward sentence like "It is hard to give a good presentation" to convey deeper meanings, like "Your talk was a mess!" One of the big puzzles in language science is how listeners work out what speakers really mean, which is a skill absolutely central to communication. However, most neuroimaging studies of language comprehension have focused on the arguably much simpler, context-independent process of understanding direct utterances. To examine the neural systems involved in getting at contextually constrained indirect meaning, we used functional magnetic resonance imaging as people listened to indirect replies in spoken dialog. Relative to direct control utterances, indirect replies engaged dorsomedial prefrontal cortex, right temporo-parietal junction and insula, as well as bilateral inferior frontal gyrus and right medial temporal gyrus. This suggests that listeners take the speaker's perspective on both cognitive (theory of mind) and affective (empathy-like) levels. In line with classic pragmatic theories, our results also indicate that currently popular "simulationist" accounts of language comprehension fail to explain how listeners understand the speaker's intended message.

Keywords: communication, fMRI, indirect replies, speaker meaning, theory of mind

Introduction

According to standard views of linguistic meaning, there is a context-invariant "sentence meaning" (coded meaning), which can be computed by retrieving relatively stable word meanings from lexical memory and by combining them in a grammatically constrained higher-order representation. Pragmatic accounts of language comprehension (e.g. Grice 1975), however, point out that the result of such lexicon- and grammar-driven sense making is actually an incomplete representation of the meaning of an utterance. Our everyday conversations seem to be full of remarks with a meaning that critically hinges on the linguistic and social context in which they are embedded. Thus, the simple phrase "The cat is on the mat" can, depending on the circumstances, be interpreted as "Can you finally get up and open the door?"; The question "Are you going to wear that tie?" is likely to result in another trip to the wardrobe; and a student hearing her teacher's statement "It's hard to give a good presentation" will probably infer that her talk might not have been a success after all. Interpreting the speaker's message (speaker meaning) requires, among other things, mechanisms for contextual disambiguation and for recovering implicit meanings that the speaker meant to convey in a particular context. It is precisely here that recent proposals in the neurobiology of language which claim that comprehension is based on sensory-motor simulation of the coded meaning (Rizzolatti and Craighero 2004) will very likely be insufficient.

This highly relevant distinction between coded meaning and speaker meaning suggests that a full account of the neurobiology of language must extend beyond systems for coding and decoding words and phrases. While they constitute a necessary point of departure, our understanding of how the brain supports natural communication will simply not be complete without grasping the neural machinery of speaker meaning comprehension. Brain imaging studies on metaphors or idioms (e.g. Mashal et al. 2007) go some way toward examining the processing of language "beyond the literal code." However, metaphors and idioms are in a way still relatively strongly tied to the code, in that they yield their "speaker meaning" independently of the particular communicative context and speaker (Holtgraves 1999). For many everyday utterances, including various forms of indirectness, the speaker meaning does critically depend on the particular context in which utterances are embedded. To study those, we need experimental paradigms in which the listener has to infer the speaker's informative intent by relying not only on the linguistic signal (as in studies on metaphor and idiom), but also on the wider discourse and social context in which the utterance serves its communicative purpose.

In the present research, we focus on the neural machinery involved in the interpretation of speaker meaning. As a test case, we contrasted direct and indirect replies—2 classes of utterances whose speaker meanings are either very similar to, or markedly different from, their coded meaning. In our study, participants listened to natural spoken dialog in which the final and critical utterance, for example, "It is hard to give a good presentation," had different meanings depending on the dialog context and the immediately preceding question. This critical utterance either served as a direct reply (to the question "How hard is it to give a good presentation?"), or as an indirect reply (to "Did you like my presentation?" or to "Will you give a presentation (rather than a poster) at the conference?"). One of the major motivations for speakers to reply indirectly in conversations is to mutually protect one another's public self or "face" (e.g. Goffman 1967; Brown and Levinson 1987; Holtgraves 1999). Half of our indirect utterances represented such emotionally charged face-saving situations, involving excuses, polite refusals, or attempts not to offend the person asking the question. The other half of the indirect replies represented more neutral situations, in which the speaker's motivation for indirectness was simply to provide more information than just a simple "no." Common to both indirect conditions was the fact that the preceding question set up a strong expectation for a yes/no answer, which was not met by a literal reading of the

indirect reply. Furthermore, and illustrated by the above example, the target utterances were identical (i.e., had the same linguistic "code") in all 3 conditions, so that any differences between the direct and indirect replies must be due to neural processes involved in speaker meaning computation.

The most influential theoretical accounts of speaker meaning interpretation stress its inferential nature (Grice 1975; Sperber and Wilson 1995; Levinson 2000; Wilson and Sperber 2004). In essence, listeners presume that speakers tailor their utterances to be optimally relevant for the present communicative situation, and any obvious departures from this relevance send the listener looking for hidden meanings. In other words, coded meaning is just a point of departure for the recovery of the actual speaker's message. In the interpretive process, listeners must also take into account information drawn from various contextual sources. These include the shared speaker–listener goals and their perspectives on the communicative situation, which has been established in the previous utterances. Hence, we expect that, relative to direct replies, interpreting indirect replies will require mentalistic inferences about the speaker's intention behind uttering a seemingly irrelevant piece of information in response to a yes/no question. At the neurobiological level, this inferential network would most likely recruit some of the regions typically involved in tasks on reasoning about the mental states of others, such as the medial frontal/prefrontal cortex, the temporo-parietal junction (TPJ), and bilateral anterior temporal lobes (Frith and Frith 2003; Saxe and Kanwisher 2003; Amodio and Frith 2006; Mitchell et al. 2006; Saxe 2006). Since the communicated meaning of indirect replies also depends on nonmentalistic inferences involving the situation model established in the prior discourse, we expect that the comprehension of indirect replies will also draw on regions in the brain that support text- and discourse-level situation model processing, beyond the classic language network (Xu et al. 2005; Ferstl et al. 2008). A brain region typically involved in such contextual anchoring is the right inferior frontal gyrus (RIFG; Menenti et al. 2009).

With respect to the 2 different types of indirect replies, we expect that they will both engage a common set of regions, since the interpretation of both requires contextual anchoring of the utterance as well as taking the speaker's perspective into account. In addition, we hypothesize that, due to their social–emotional connotations, listening to face-saving indirect replies will engage socio-cognitive and/or affect-related brain structures, such as the amygdala, the anterior cingulate cortex (ACC; e.g. Dalgleish 2004), insula (Fan et al. 2011), or the anterior/inferior temporal lobe (Binder and Desai 2011).

An influential alternative to the inferential view of language comprehension is inspired by the discovery of the mirror neurons and often referred to as the "simulationist" view. The "simulationist" view states that comprehension does not require any inferential steps, but can work by virtue of simulation, or automatic sensorimotor processes, creating a common semantic link between the speaker and the addressee (Rizzolatti and Craighero 2004). Critically, the implicit assumption here is that it is the coded meaning of the utterance that is simulated, which is in contrast with the above view that gives a central place to linguistic and socio-cognitive inferences. The simulation is thought to be accomplished by means of the human mirror-neuron system, presumably located in the occipital, temporal, and parietal visual regions, as well as the inferior parietal lobule, the precentral gyrus, and area 44 of the inferior frontal gyrus (e.g. Rizzolatti et al. 1996; Iacoboni 1999; Buccino et al. 2001; Iacoboni et al. 2001).

In the domain of understanding language, the putative mirror-neuron system has been argued to be implicated in, for example, action word comprehension (see e.g. Pulvermüller and Fadiga 2010), but also when people interpreted the message conveyed by communicative gestures during a game of charades (Schippers et al. 2009). However, this proposal has never been tested at such a high level of meaning processing, in the computation of speaker meaning. Since deriving speaker meaning is such an essential aspect of human communication, it is crucial that alternative views on the necessary neurobiological infrastructure for language comprehension no longer ignore neuropragmatic aspects of language. Here, we focus on the neural machinery involved in the recovery of speaker meaning. As part of that, we will explore whether this interpretation process is inferential in nature, or whether understanding of communicative messages takes place via inference-free simulation of coded meaning of the speaker's utterance. In our study, participants were listening to dialogs between 2 people while their blood oxygen level-dependent (BOLD) signals were acquired in the MR scanner. In one condition, the answer that the second person provided was a direct reply to the question of the first person. In the other conditions, the same answer was an indirect reply to another question by the first person. We only contrasted the activation to the same answers in the different reply modes.

Results

The crucial comparison for the issue at stake is the one between same sentences in their role as indirect versus direct replies. Listening to replies whose meaning was indirect in contrast to replies with a more direct meaning activated a large frontal and medial prefrontal network, including bilateral superior medial frontal gyrus, right supplementary motor area (SMA), and parts of the inferior frontal gyrus (pars orbitalis and pars triangularis) bilaterally, extending into the insula in the left hemisphere (see Fig. 1; for exact coordinates, see Table 1). In addition, the right TPJ and the right middle temporal gyrus showed an increased activation. To disentangle the relative contribution of the 2 types of indirect replies, we performed several additional analyses. Interpreting the face-saving replies activated a network of regions largely overlapping with the the overall comparison between indirect and direct replies. For this comparison, medial frontal cortex (MFC) activation included the left ACC, and an additional activation cluster was present in the right superior temporal gyrus (STG; Supplementary Material, Supplementary Fig. 1B). When we excluded the face-saving indirect replies, still a significant subset of the regions from the pooled comparison remained active: The right TPJ, left insula, as well as bilateral inferior frontal gyrus: Pars orbitalis in the right hemisphere and pars triangularis in the left hemisphere. An additional activation was seen in the left temporal pole (Supplementary Fig. 1A).

Finally, a direct comparison of processing triggered by the face-saving indirect replies in comparison to the informative indirect replies revealed a number of clusters in the right hemisphere, of which the largest ones were in STG and the ACC. Once again, there was a cluster spanning the RIFG and anterior insula, although this time it was only marginally significant, at
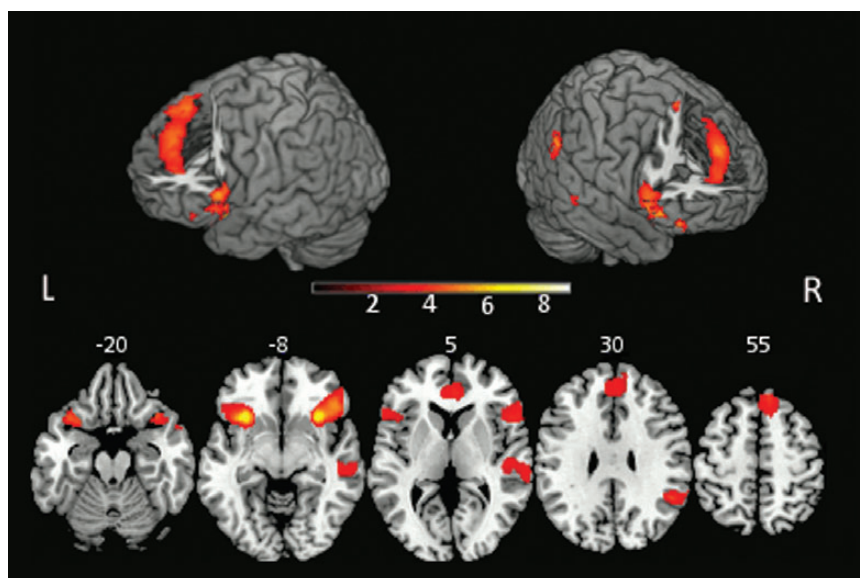
Figure 1. Activations for the main effect of indirectness. Significant effects are displayed on cortical renderings and on axial slices (z coordinate levels in millimeters).

$P < 0.10$ (Supplementary Fig. 2). Results and figures from the conjunction analysis of the 2 indirect effects are presented in Supplementary Material (Supplementary Figs 3 and 4).

Discussion

The main goal of this study was to identify brain regions involved in language comprehension at a level of representation typically overlooked in the neurobiology of language research: The intended meaning a speaker wants to communicate with a specific utterance, also known as speaker meaning (Grice 1957). Participants listened to utterances that were identical at the word and sentence level, but they were used to express different informative intentions. This allowed us to isolate processing related to speaker meaning interpretation beyond the retrieval of individual word meanings, sentence-level semantic composition, or even low-level pragmatic enrichment, such as fixing the referents of pronouns. We have shown that deriving the speaker's communicative intention depends on several brain regions previously implicated in mentalizing and empathy (MFC, right TPJ, and the anterior insula) as well as in discourse-level language processing (bilateral prefrontal cortex and right temporal regions). Moreover, we have shown that when the speaker meaning has affective implications, a number of right-lateralized regions get involved. These regions are previously implicated not only in affective and social-cognitive processing (insula and ACC), but also in building and maintaining a coherent representation of what is going on in the discourse (IFG and STG).

In the indirectness effect (comprising both types of indirect replies against a direct-reply baseline), there were activations in the MFC extending into the right anterior part of the SMA, and in the right TPJ, a pattern typical for tasks that involve higher-order, theory-of-mind (ToM)-like mentalizing (Amodio and Frith 2006; Mitchell et al. 2006; Saxe et al. 2006). Although the exact role of all the individual ToM regions is not yet clearly established, both MFC and right TPJ constitute core regions activated across various input modalities (such as in cartoons or auditory presented stories) and in both verbal and

nonverbal tasks (Carrington and Bailey 2009) in ToM research. The most specific hypothesis about the role of the right TPJ in the mentalizing network is that it is implicated in mental-state reasoning, that is, thinking about other people's beliefs, emotions, and desires (Saxe 2010). Activation in the right TPJ has been also shown to correlate with autistic spectrum disorder syndrome severity in a self-other mental-state reasoning task (Lombardo et al. 2011).

An MFC cortex is a large cortical region with a variety of roles characteristic of social cognition in general, beyond ToM processing (Amodio and Frith 2006; Saxe and Powell 2006). Based on a meta-analysis of task-related activations from this region, Amodio and Frith (2006) proposed a division of the MFC into 3 distinct functional and anatomical regions with different connections to the rest of the brain. The peaks of our activation, although fairly close to each other, fall in the anterior and posterior rostral divisions, which are associated with complex socio-cognitive processes such as mentalizing or thinking about the intentions of others (such as communicative intentions and right anterior MFC) or oneself (right posterior MFC). Interestingly, the involvement of these regions is not exclusive to ToM tasks, but is consistently observed in the narrative comprehension literature (e.g. Mason and Just 2009; Mar 2011). This is not surprising, as it is likely that the motivations, goals, and desires of fictional characters are accessed in a similar manner as with real-life protagonists (Mar and Oatley 2008). In fact, an influential model from the discourse processing literature (Mason and Just 2009) ascribes the dorsomedial part of the frontal cortex and the right TPJ a functional role as a protagonist perspective network, which generates expectations about how the protagonists of stories will act based on understanding their intentions.

In the context of speaker meaning interpretation, the fact that we found activation in these 2 brain regions typically involved in social cognition suggests that listeners engage in nonlinguistic perspective taking in order to fully comprehend the meaning of the indirect replies. Just as theoretical accounts suggest (e.g. Grice 1975), getting at the speaker's intended message means that the listener considers not only the

Table 1
Activations for contrasts of interest thresholded at 0.001

Anatomical region	Coordinates of local maxima			BA	Cluster size	P-value (cluster-level FWE corrected)
	x	y	z			
Pooled indirect replies > direct replies						
R supplementary motor area	14	24	58	8	1641	<0.001
L middle frontal medial gyrus	−4	42	28	32		
R middle frontal medial gyrus	8	36	50	8		
R inferior frontal gyrus	34	22	−12	47/49	1592	<0.001
R inferior frontal gyrus	44	30	−4	47		
R inferior frontal gyrus	58	22	4	47		
L anterior insula	−30	16	−14	15	1381	<0.001
L inferior frontal gyrus	−58	22	10	47		
L inferior frontal gyrus	−52	38	−6	47		
R temporo-parietal junction	48	−50	32	39/40	375	0.003
R middle temporal gyrus	52	−28	−6	21	248	0.022
	62	−30	−2	21		
Indirect informative replies > direct replies						
L temporal pole	−36	16	−20	38	472	0.001
L inferior frontal gyrus	−52	20	6	47		
L anterior insula	−28	20	−4	13/15		
R temporo-parietal junction	48	−50	32	39/40	281	0.011
R inferior frontal gyrus (pars orbitalis)	38	22	−12	47/49	189	0.049
Indirect face-saving replies > direct replies						
R anterior cingulate cortex	6	44	16	24/32	2137	<0.001
R supplementary motor area	14	24	58	8		
R superior medial cortex	10	34	56	8		
R anterior insula	34	20	−12	15	1958	<0.001
R inferior frontal gyrus (pars orbitalis)	48	28	−4	47		
R inferior frontal gyrus	60	20	16	45		
L anterior insula	−32	16	−14	15	1295	<0.001
L inferior frontal gyrus (pars triangularis)	−58	24	8	45		
R superior temporal gyrus	58	−22	2	22	763	<0.001
	64	−28	0	21/22		
R middle temporal gyrus	54	−28	−4	21		
R temporo-parietal junction	50	−50	32	39/40	247	0.030
Indirect face-saving replies > indirect informative replies						
R superior temporal gyrus	56	−20	4	41/42/22	1011	<0.001
	50	−26	8	41/42/22		
	56	−6	−8	22		
R anterior cingulate cortex	6	48	18	32	414	0.002
R inferior frontal gyrus (pars orbitalis)	44	26	−14	47	166	0.091
R anterior insula	32	20	−16	15		

Note: Cluster P-values are corrected for multiple nonindependent comparisons. All reported coordinates are in the MNI space. BA, Brodmann area; FWE, family-wise error.

meanings of her words, but also what was her motivation and what goal she wanted to achieve when she uttered these words in both the specific linguistic and social context.

The general comparison of direct and indirect replies also engaged the left insula, a region known to be involved in empathy and affective processing (Singer and Lamm 2009; Berntson et al. 2011). One plausible explanation of anterior insula involvement in deriving speaker meaning is that it provides a low-level form of perspective taking (see also Ruby and Decety 2001), the outcome of which then might be "relayed" to higher-level mentalizing processes. This interpretation of insula involvement is supported by a recent meta-analysis of studies involving a wide variety of empathy-invoking stimuli and tasks (Fan et al. 2011), which found a division of labor in the anterior insula based on laterality, with the left insula implicated in both affective–perceptual and cognitive–evaluative forms of empathy. Taken together, this suggests that speaker meaning interpretation requires 2 types of nonlinguistic perspective taking: A more reasoning-based perspective taking ("What does the protagonist think?") and a more experiential, affective appreciation of "how does it feel to be the protagonist." [The "face-saving effect" revealed in the contrast between face-saving and informative indirect replies is further discussed in Supplementary Material (Supplementary Fig. 6).]

Also involved in recovering the meaning of indirect replies were parts of the perisylvian language network in the left inferior frontal gyrus (Brodmann area 45 [BA45] and 47) and their right-hemispheric homologs. The left IFG plays a prominent role in language processing (Price 2000), from sentence-level processes such as semantic unification of lexical information (Hagoort et al. 2009) to linking causally related sentences when reading texts (Kuperberg et al. 2006). Text comprehension research suggests that regions within the bilateral IFG might support the semantic selection of inferential information (Mason and Just 2011). In addition, the right IFG seems to be particularly related to constructing a situation model based on linguistic and nonlinguistic inputs (Menenti et al. 2009). This interpretation of IFG function is consistent with the fact that the meaning of indirect replies is crucially dependent on the linguistic and social details of the context they are embedded in.

Simulation or Inference?

One influential view on how language comprehension is implemented in the brain is the simulationist view, endorsing the existence of direct, automatic, cognitively unmediated sensorimotor resonance processes that establish common semantic links between the speaker and listener (Rizzolatti and Craighero 2004). On the neural level, these are supposedly implemented in the brain regions that contain mirror neurons or have mirror properties. A crucial assumption is that listeners re-enact the coded meaning of utterances, and there is no place for linguistic or nonlinguistic (such as mentalizing) inferential processes in this model. Although this approach might be able to explain language comprehension at a single word level, our results show that such reasoning "cannot be the whole story" if we consider higher levels of meaning.

Only 2 brain regions from the comparison between direct and indirect replies, the insula and SMA, can be considered part of the simulation network, and there are alternative accounts for both of them (Picard and Strick 2001; Saxe 2005, 2009; Decety 2010). Even if we accept their role as "mirroring" components of our speaker meaning interpretation network, they clearly need the support of other language and mentalizing-related regions, which is not consistent with a "cognitively unmediated" explanation of language comprehension. Thus, we conclude that language comprehension in its most typical niche—in rich social contexts—goes beyond simulating coded meanings of words and sentences.

While our results speak against a simple sensorimotor, or mirror-neuron type of simulation, we cannot rule out that

language comprehension depends partly on a more complex form of simulation. There are currently simulationist positions in which simulation is based on the recognition and assessing of the intentions of others (e.g. Goldman 2006, 2009), as well as a recent model of language production and comprehension where both of these processes depend on simulating the speaker's intentions (Pickering and Garrod forthcoming). What these richer simulationist accounts have in common is that even though they endorse simple mirroring as a mechanism playing part in action/language comprehension, it is not the sole mechanism of comprehension.

In conclusion, we have presented evidence suggesting that meaning interpretation in communication is fundamentally inferential in nature, with a critical role in arriving at the intended meaning of the speaker's message played by linguistic and mentalizing inferences. These results are in direct opposition to the view that comprehension can be accomplished by direct, cognitively unmediated "simulation" of the coded meaning of speakers' utterances. Instead, we suggest that listeners take the speaker's perspective at both cognitive (ToM) and affective (empathy-like) levels.

Experimental Procedures

Participants
Twenty-eight native speakers of Dutch participated in the experiment (5 males, mean age 21.2 years, SD = 2.67). Three additional subjects were excluded from the analysis because of excessive head movement during scanning. All participants were right-handed and had no history of neurological impairment or head injury. They all signed an informed consent form and received payment or course credits.

Stimulus Material
We created 90 critical utterances that were preceded by 3 different types of context, making up 270 experimental items in total. There were 3 experimental conditions in the study. Depending on the preceding context, each critical utterance could be interpreted as either a direct reply (condition 1) or an indirect reply (conditions 2 and 3). While one of the indirect conditions was purely informative, the other involved a socioemotional aspect, as the reason for indirectness was to "save one's face" (as in excuses or polite refusals). Thus, the speaker meaning was either explicitly stated and largely corresponded to the sentence-level meaning of the critical utterance (direct), or, in both indirect conditions (informative, face-saving), the speaker's message was implicit and a pragmatic inference was necessary to recover it. A small number of the indirect replies were adapted from Holtgraves (1999).

The structure of each item was as follows: A lead-in story set up the relevant context, introducing the 2 lead characters and any necessary background (e.g. where they are, what they are doing, what their goals are). After that, the characters held a short 4-turn dialog culminating in the critical question-reply pair. Across the 3 conditions, the critical utterance (reply) was always the same. Thus, we had 2 types of context: A wider background (lead-in story) as well as the immediate context (critical question). Table 2 provides an example of the stimulus materials (for a detailed description of how the stimuli were constructed, see Supplementary Material).

Table 2

Examples of the 3 different types of replies, preceded by their respective context stories, translated into English

Direct reply
John needs to earn some extra course points. One of the possibilities is to attend a student conference. He has never been to a conference before, and he has to decide whether he wants to present a poster, or give a 15-min oral presentation. He is talking to his friend Robert, who has more experience with conferences. John knows that Robert will be realistic about how much work it takes to prepare for a conference.
J: How is it to prepare a poster?
R: A nice poster is not so easy to prepare.
J: And how about a presentation?
R: *__It's hard to give a good presentation__*.

Indirect informative reply
John and Robert are following a course in Philosophy. It is almost the end of the semester. The lecturer has announced that they can either write a paper, or give a presentation about a philosopher of their choice. Both John and Robert are ambitious students and want to get good grades. They know that they want to talk about postmodern philosophers, but they are not yet sure about the format. They are discussing their possibilities.
J: I think that I will rather write a paper.
R: I agree, you are a very good writer.
J: Will you choose a presentation?
R: *__It's hard to give a good presentation__*.

Indirect face-saving reply
John and Robert are following a course in Philosophy. It is the last lesson of the semester, and everybody has to turn in their assignments. Some people have written a paper, and others have given a presentation about a philosopher of their choice. John has chosen the latter. When the lesson is over, he is talking to Robert.
J: I'm relieved it's over!
R: Yes, the lecturer was really strict.
J: Did you find my presentation convincing?
R: *__It's hard to give a good presentation__*.

The target utterance is always the final one (in bold italics).

In addition, there were 55 filler items. The purpose of the filler items was 2-fold: First, approximately two-thirds of the final utterances of the filler dialogs were more explicit than the critical utterances, containing yes/no and similar expressions. Secondly, after 50 of the filler items, participants had to answer a visually presented true/false statement with a button press. A correct reply required them to process the filler item for its implicit meaning. No other task demands were imposed. Two filler items were presented as example items before the actual experiment, and one filler item was used to adjust the sound level for each participant before each scanning session.

All items were presented auditorily. The lead-in stories were narrated by a female Dutch speaker. The dialogs were recorded by 80 male and female Dutch native speakers. We chose the speakers with respect to the age and sex of the dialogs' protagonists, and each speaker recorded 3 dialogs on average. Each dialog was recorded several times in order to choose the best version. The recordings were edited in Praat (Boersma 2001), and 2 native Dutch speakers then jointly chose the best recording for each of the 3 conditions.

Procedure
The study consisted of 2 sessions lasting approximately 1.5 h each, on 2 different days. One session comprised 2 experimental blocks, the other 3 experimental blocks, with counterbalanced order of presentation. Each experimental block consisted of 18 critical items and 11 filler items. There was a short break after each block.

Participants received written instructions before the experiment, asking them to listen to the stories and dialogs. They were asked to pay special attention to "what the protagonists really wanted to say" with their final utterances and were reminded that this "speaker's message" is, in light of the context,

sometimes similar and at other times different than the actual words the protagonists are using. The instructions also contained 3 example items.

Each stimulus was preceded by a fixation cross for 2 s. The lead-in story was then presented in stereo via headphones. After the last sentence of the story, the names of the 2 protagonists depicted in the story were displayed on the screen for another 2 s, one on the left and another on the right. The left/right assignment corresponded to the direction from which the participant heard the particular speaker during the dialog, to ease protagonist identification. During the entire dialog, a fixation cross was in the middle of the screen and remained there for another 4 s after the end of target utterance.

Each participant heard 145 stories and dialogs in total. Because each target critical utterance was presented only once to a single participant, we constructed 3 stimulus lists.

We matched the 3 conditions within each list as closely as possible on the following characteristics: Length of each target utterance (in seconds) and length of the preceding context (in words), lexical frequencies of the content words in the critical utterances based on frequency counts from the Spoken Dutch Corpus [corpus gesproken Nederlands e.g. Oostdijk 2000], semantic similarity of the context stories and dialogs up to the target utterance, and finally the amount of direct semantic priming (repetition of the same content words) from the lead-in story and from the critical question. Each list was pseudorandomized, with no more than 2 items from the same condition appearing after each other.

The block order, Left/Right assignment of the speakers, and TRUE/FALSE button assignment in the task were counterbalanced across participants.

Task

On 50 of the filler items, participants had to answer a true/false statement. The statement could only be answered correctly if participants paid attention to the speakers' message, which was mostly not explicitly stated.

The statements were presented visually after the last sentence of the dialog and stayed on the screen until the participants had responded by pressing the left or right button with their left or right index finger.

fMRI Data Acquisition

Participants were scanned in an ascending fashion with a Siemens 3-T Tim-Trio MRI-scanner, using a 8-channel surface coil. The repetition time (TR) was 2.4 s and each volume consisted of 35 slices of 3-mm thickness with a 17% slice gap. The voxel size was $3.5 \times 3.5 \times 3$ mm^3, and the field of view was 224 mm. Functional scans were acquired at echo time (TE) = 30 ms. Flip angle was 80°. A whole-brain high-resolution structural T_1-weigthed GRAPPA sequence was performed to characterize participants' anatomy (TR = 2300 ms, TE = 3.03 ms, 192 slices with voxel size of 1 mm^3, field of view = 256).

fMRI Data Analysis

The functional magnetic resonace imaging (fMRI) data were preprocessed and analyzed using Statistical Parametric Mapping (SPM5, fil.ion.ucl.ac.uk/spm/). The first 5 images in each session were discarded to prevent a transient nonsaturation effect from affecting the analysis. The functional echo-planar imaging-BOLD images were then realigned and

slice-time corrected. The resulting functional images were coregistered to the participants' anatomical volume based on the subject-mean functional image, normalized to MNI space, and spatially smoothed using a 3-dimensional isotropic Gaussian smoothing kernel (full-width half-maximum = 8 mm). A temporal high-pass filter was applied with a cycle cutoff at 128 s.

In the first-level linear model, we modeled the onsets and durations of the 3 types of the target utterances (direct, indirect face-saving, and indirect informative), which were defined as the entire conversational turn, including the short preutterance silence. Each of the conditions included 30 trials. We also modeled the onset and duration of the visually presented 2 s fixation cross before each experimental item (baseline), as well as the 4-s fixation cross after the end of the target utterance. The regressors were convolved with a canonical hemodynamic response function, and the realignment parameters were included in the model to correct for subject movement during scanning. Subsequently, various images were defined for each participant and used in the second-level random effects analysis.

In the second-level random effects analysis, we used the contrast images of interest in a repeated-measures analysis of variance. The cluster size was used as the test statistic, and only clusters significant at $P < 0.05$ corrected for multiple nonindependent comparisons are reported. The initial threshold was 0.001 at the voxel level.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

References

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

Berntson GG, Norman GJ, Bechara A, Bruss J, Tranel D, Cacioppo JT. 2011. The insula and evaluative processes. Psychol Sci. 22:80–86.

Binder JR, Desai RH. 2011. The neurobiology of semantic memory. Trends Cogn Sci. 15:527–536.

Boersma P. 2001. Praat, a system for doing phonetics by computer. Glot International. 5(9/10):341–345.

Brown P, Levinson SC. 1987. Politeness: some universals in language usage. Cambridge (UK): Cambridge University Press.

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ. 2001. Action observation

activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur J Neurosci. 13:400–404.

Carrington SJ, Bailey AJ. 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. Hum Brain Mapp. 30:2313–2335.

Dalgleish T. 2004. The emotional brain. Nat Rev Neurosci. 5:583–589.

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

Fan Y, Duncan NW, De Greck M, Northoff G. 2011. Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. Neurosci Biobehavior Rev. 35:903–911.

Ferstl E, Neumann J, Bogler C, Von Cramon DY. 2008. The extended language network: a metaanalysis of neuroimaging studies on text comprehension. Hum Brain Mapp. 29:581–593.

Frith U, Frith CD. 2003. Development and neurophysiology of mentalizing. Philos Trans R Soc B Biol Sci. 358:459–473.

Goffman E. 1967. Interaction ritual: essays on face-to-face interaction. New York (NY): Pantheon.

Goldman AI. 2009. Précis of simulating minds: the philosophy, psychology, and neuroscience of mindreading. Philos Studies. 144:431–434.

Goldman AI. 2006. Simulating minds. The philosophy, psychology, and neuroscience of mindreading. New York: Oxford University Press.

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

Grice HP. 1957. Meaning. Philos Rev. 66:377–388.

Hagoort P, Baggio G, Willems R. 2009. Semantic unification. In: Gazzaniga MS, editor. The cognitive neurosciences. Cambridge (MA): MIT Press. p. 819–836.

Holtgraves T. 1999. Comprehending indirect replies: when and how are their conveyed meanings activated? J Mem Lang. 41:519–540.

Iacoboni M. 1999. Cortical mechanisms of human imitation. Science. 286:2526–2528.

Iacoboni M, Koski LM, Brass M, Bekkering H, Woods RP, Dubeau M-C, Mazziotta JC, Rizzolatti G. 2001. Reafferent copies of imitated actions in the right superior temporal cortex. Proc Natl Acad Sci USA. 98:13995–13999.

Kuperberg GR, Lakshmanan BM, Caplan DN, Holcomb PJ. 2006. Making sense of discourse: an fMRI study of causal inferencing across sentences. NeuroImage. 33:343–361.

Levinson SC. 2000. Presumptive meanings. Cambridge (MA): MIT Press.

Lombardo MV, Chakrabarti B, Bullmore ET, Baron-Cohen S. 2011. Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. NeuroImage. 56:1832–1838.

Mar RA. 2011. The neural bases of social cognition and story comprehension. Ann Rev Psychol. 62:103–134.

Mar RA, Oatley K. 2008. The function of fiction is the abstraction and simulation of social experience. Perspect Psychol Sci. 3:173–192.

Mashal N, Faust M, Hendler T, Jung-Beeman M. 2007. An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. Brain Lang. 100:115–126.

Mason RA, Just MA. 2011. Differentiable cortical networks for inferences concerning people's intentions versus physical causality. Hum Brain Mapp. 32:313–329.

Mason RA, Just MA. 2009. The role of the theory-of-mind cortical network in the comprehension of narratives. Lang Linguist Compass. 3:157–174.

Menenti L, Petersson KM, Scheeringa R, Hagoort P. 2009. When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. J Cogn Neurosci. 21:2358–2368.

Mitchell JP, Macrae CN, Banaji MR. 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron. 50:655–663.

Oostdijk N. 2000. Het Corpus Gesproken Nederlands. Nederlandse taalkunde. 5:280–284.

Picard N, Strick PL. 2001. Imaging the premotor areas. Curr Opin Neurobiol. 11:663–672.

Pickering MJ, Garrod S. forthcoming. An integrated theory of language production and comprehension. Behav Brain Sci.

Price CJ. 2000. The anatomy of language: contributions from functional neuroimaging. J Anat. 197:335–359.

Pulvermüller F, Fadiga L. 2010. Active perception: sensorimotor circuits as a cortical basis for language. Nature Rev Neurosci. 11:351–360.

Rizzolatti G, Craighero L. 2004. The mirror-neuron system. Annu Rev Neurosci. 27:169–192.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L. 1996. Premotor cortex and the recognition of motor actions. Brain Res. 3:131–141.

Ruby P, Decety J. 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. Nat Neurosci. 4:546–550.

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

Saxe R. 2009. The neural evidence for simulation is weaker than I think you think it is. Philos Studies. 144:447–456.

Saxe R. 2010. The right temporo-parietal junction: a specific brain region for thinking about thoughts. In: Leslie A, German T, editors. Handbook of theory of mind. 1st ed. Philadelphia, PA: Psychology Press.

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. 19:1835–1842.

Saxe R, Moran JM, Scholz J, Gabrieli J. 2006. Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. Soc Cogn Affect Neurosci. 1:229–234.

Saxe R, Powell LJ. 2006. It's the thought that counts: specific brain regions for one component of theory of mind. Psychol Sci. 17:692–699.

Schippers MB, Gazzola V, Goebel R, Keysers C. 2009. Playing charades in the fMRI: are mirror and/or mentalizing areas involved in gestural communication? PloS ONE. 4:e6801.

Singer T, Lamm C. 2009. The social neuroscience of empathy. Ann N Y Acad Sci. 1156:81–96.

Sperber D, Wilson D. 1995. Relevance: communication and cognition. 2nd ed. Oxford: Blackwell.

Wilson D, Sperber D. 2004. Relevance theory. In: Horn LR, Ward G, editor. The handbook of pragmatics. Oxford: Blackwell. p. 607–632.

Xu J, Kemeny S, Park G, Frattali C, Braun A. 2005. Language in context: emergent features of word, sentence, and narrative comprehension. NeuroImage. 25:1002–1015.