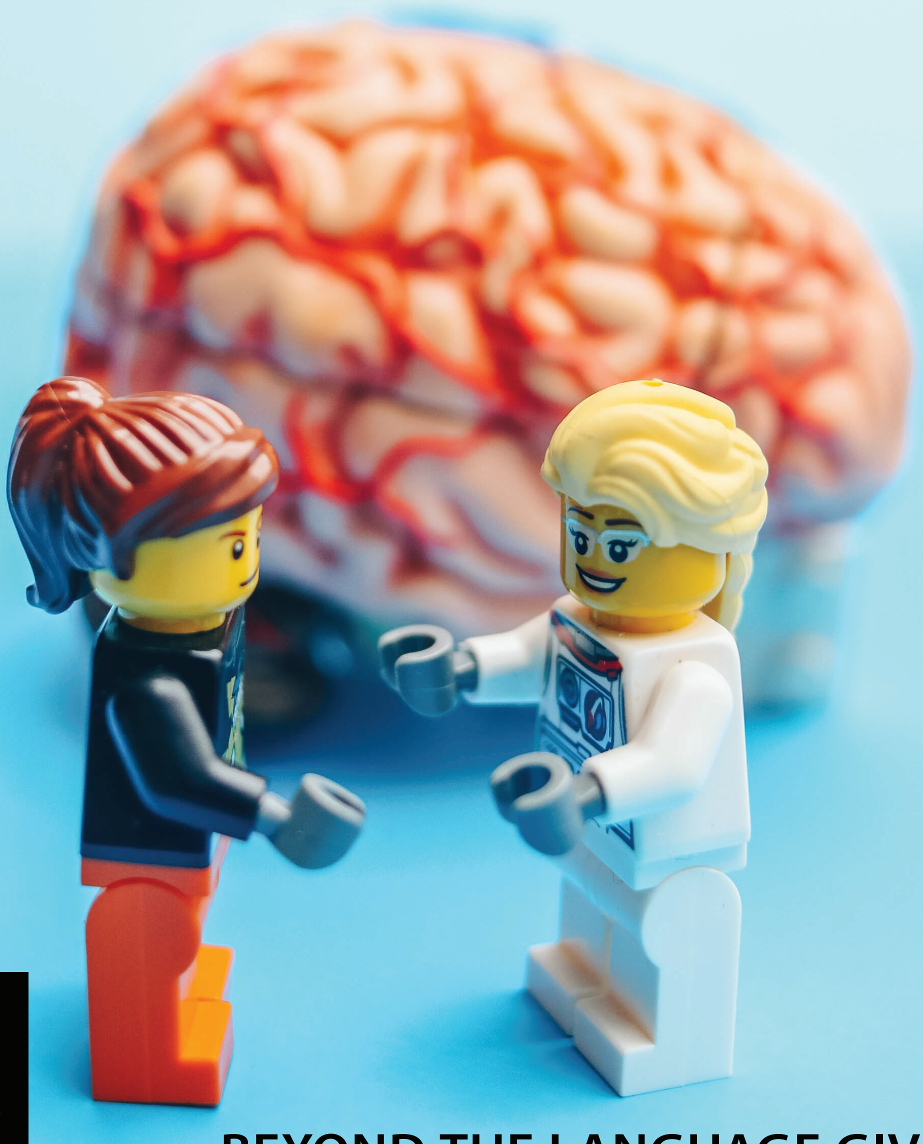
JANA BAŠNÁKOVÁ



BEYOND THE LANGUAGE GIVEN

THE NEUROBIOLOGICAL INFRASTRUCTURE FOR PRAGMATIC INFERENCING

Max Planck Institute for Psycholinguistics
Series

Beyond the language given: the neurobiological infrastructure for pragmatic inferencing

Jana Bašnáková

Beyond the language given: the neurobiological infrastructure for pragmatic inferencing

PhD Thesis, Max Planck Institute for Psycholinguistics

Cover artwork by Jana Bašnáková & Miroslav Kolárik

Printed and bound by Ipskamp Drukkers bv

Beyond the language given: the neurobiological infrastructure for pragmatic inferencing

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,

volgens besluit van het college van decanen

in het openbaar te verdedigen op woensdag 23 januari 2019

om 10.30 uur precies

door

Jana Bašnáková

geboren op 29 juli 1979

te Bratislava, Slowakije

Promotors: Prof. dr. Peter Hagoort

Prof. dr. Jos van Berkum

Manuscriptcommissie: Prof. dr. Stephen C. Levinson

Prof. dr. Ivan Toni

Dr. Judith Holler

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

1. Introduction

On the surface, understanding spoken language is about decoding signals that are first neatly "packaged" by the speaker, and then transferred to the addressees' ears as soundwaves. In the addressees' brain, the meaning corresponding to these soundwaves is retrieved from the mental lexicon and, with the help of learned grammatical rules, assembled into a meaningful idea.

The so-called "code model" of communication is built on two ideas: firstly, the conduit metaphor (Reddy, 1979), stating that information from the speaker can be wrapped up in words and sent along a specific channel for the listener to unwrap; and secondly, information theory (Shannon, 1948) about signals as strings of information transmitted along a noisy channel for the listener to reconstruct them.

However, only a few examples of how people actually use language in everyday interactions suffice to show that such a "transport" perspective is overly simplistic (Sperber & Wilson, 1995). When a hospital nurse informs a colleague that "*the appendix in room 4 still needs her dinner*", there is mutual understanding that it is not an actual appendix that is hungry, but rather a patient hospitalized with appendicitis. Saying "*I am cold*" on a date could be meant as neutral information, but also as an invitation to snuggle up. A pointing gesture towards the cat's empty bowl can be intended as a reminder for a roommate who has trouble keeping her feeding schedule in mind. And hearing "*It's hard to give a good presentation,*" (Holtgraves, 1998) from a well-meaning colleague could be simply a statement about the hardships of using Powerpoint, but also rather compassionate feedback about your less-than-ideal performance at a conference.

A great deal of meaning in conversations does not lie within the actual words and the way they are arranged into sentences according to the rules of syntax. The intended message – or "speaker meaning" -- is often far richer than what the speaker explicitly utters. In order for communication to be successful, listeners are not passive "receivers", but actually do a lot of work. They must, for instance, draw on a variety of contextual information. Some of it is necessary to decide which of the multiple possible senses to choose for ambiguous words, such as whether "*hard*" is meant as "*not easily broken*" or "*difficult*". Other details are inferred and filled in from what the addressee already knows: for example, that it is often rude to convey negative information in a direct way because it can hurt someone's feelings (hence, one would not typically say "*Your presentation was awful!*").

In this thesis, I investigate the neural correlates of interpreting such speaker meaning, through studying indirect replies. Indirect replies are a good window into the various kinds of "extra" work listeners do in making sense of what their conversational partners mean, for two reasons. First of all, as the "presentation" example showed, what the addressee hears or reads "on the surface" is often vastly different from what the speaker intends to convey, prompting the addressee to bridge the gap between what is said and what is actually meant through various inferences. In this thesis, I am zooming in on the nature of the pragmatic inferences that enable addressees to grasp speaker meaning.

More specifically, I want to find out whether uncovering the implicit meanings of an indirect conversational remark engages brain regions different from the "core", left-lateralized fronto-temporal network that is typically active in "ordinary", non-implicit language comprehension tasks (Hagoort & Indefrey, 2014). What regions contribute to the interpretation of meanings that go beyond words and sentences? There are several candidates, including right-hemisphere homologues of the core language areas, as well as networks implicated in non-language processes, for example reasoning about the beliefs, emotions and desires of others.

Secondly, indirect utterances also reflect the complicated interpersonal work that conversational partners routinely do to facilitate smooth interactions and to achieve their goals. Indirectness is typically used in situations where the speaker wants to avoid losing her face, or help to protect the face of the addressee (Brown & Levinson, 1987; Goffman, 1959), or where she wants to deliberately convey a potentially risky message in an ambiguous way so that it can be later denied, if necessary (Pinker, Nowak, & Lee, 2008). In this sense, indirectness also gives us the opportunity to investigate the interface between linguistic and social processes. Given that the interpersonal work we do in communication significantly contributes to the success with which we navigate our social worlds (Brown & Levinson, 1987; Holtgraves, 2002), we can explore whether changing the addressee's interpersonal involvement will also affect her linguistic processing. What are the consequences of hearing an indirect remark as a direct addressee, in contrast to being just an overhearer? Will it affect neural networks implicated in linguistic processing, or rather regions regulating social interaction and emotions, or both?

Words are not the only cues to interpretation. A lot is achieved through non-verbal means – how we gesture, shake our head, how we time our response. These are unconventional communicative devices without "set" intrinsic meaning – and yet, as apt communicators, we can make sense of them and interpret them quickly and efficiently. In this thesis, I also investigate indirect meanings expressed via simple pointing gestures (Tomasello, 2008). Are the pragmatic inferences behind gestural indirectness identical with those in verbal indirectness? Will pointing gestures used to express indirect meanings engage language-related brain areas, regions regulating social interaction, and/or hand-action related regions?

Investigating language used for communicative purposes, with its implicit meanings and interpersonal aspects, has traditionally been difficult for psycholinguists and cognitive neuroscientists. In real life, communication is like a fleeting, complicated dance: there are many simultaneous moves at various levels, with conversational partners quickly reacting and adjusting to each other's signals. It becomes a huge challenge accounting for such complexity in experiments. The powerful technology we now have to look at the inner workings of the brain – functional Magnetic Resonance Imaging (fMRI) – introduces additional methodological hurdles. Participants have to lie still in a small, confined space, moving no more than a few millimeters in order not to cause unwanted artifacts. Because the signal emitted by the cognitive processes that we study is far smaller than noise caused by other sources, we have to use a lot of repetition of the specific types of communicative acts we are looking at. In this thesis, I attempt to find a balance between the richness of communicative situations and strict experimental control. To investigate indirectness comprehension, I developed three dialog paradigms in which it is possible to isolate the key contrasts in a way that the participants find indirect replies fairly natural.

2. Levels of meaning and intentions in communication

In this thesis, I talk a lot about meaning. But – what does "mean" mean (Noveck & Reboul, 2008)? Imagine an interaction between a husband and his wife getting ready for a dinner party. The man dresses himself and finishes with an extravagant red tie. He then asks his wife: "*Do you like this?*" She looks at him and says "*I think that this is going to be a pretty formal event, darling*".

In the wife's reply, we can distinguish at least 3 different levels of meaning (e.g. Clark, 1996, 1997). The first level, *sentence meaning,* refers to the invariable, timeless, "dictionary" meaning of an utterance, based on conventionalized sign meanings. The standard case of sentence meaning involves word meanings combined in a declarative sentence according to the rules of grammar (cf. Jackendoff, 2007). However, there can be special cases like conventionalized idioms (*Kick the bucket*) or conventional metaphors (*Lawyers are sharks*), where sentence meaning refers to the entire construction. Sentence meanings can be established by theorists for the *type*, but will also in some way be computed by language users for the *token* (i.e. a particular instance of the type), using their knowledge of conventionalized signs and their combination.

Sentence meaning is a representation where several pieces of information are not yet specified. For instance, the referents of deictic expressions (such as *this*) or pronouns (*I*) are not fixed, as well as meanings of ambiguous words (e.g. *pretty*) are not determined. After filling in these parameters, we can talk about the second level, or *utterance meaning. Utterance meaning* is the context-dependent meaning of a sentence uttered by a particular speaker in a particular context (i.e. the use of a sentence as a *token*). Getting from sentence to utterance meaning involves sense disambiguation (here, "*pretty*" refers to "*very*"), establishing references ("*I*" refers to the wife), and other enrichment. Utterance meaning is clear for the speaker, and must be recovered by the addressee on the basis of context.

Sense disambiguation at this level also includes the addressees choosing between the literal (compositional) or figurative reading of idioms; it is also here where they must arrive at the relevant features of metaphorically used words. Attending to context in constructing utterance meaning must involve some level of perspective taking (which of the several senses of the word does the speaker mean?).

The third level – *speaker meaning* - expresses what the speaker really wants to convey to the addressee[1]. Speaker meaning is the contextualized social intention of a particular speaker

[1] Another way to refer to the distinction between utterance and speaker meaning is "what is said" (utterance meaning) and "what is implicated" (speaker meaning) (see e.g. Garrett & Harnish, 2007).

that is pursued by uttering a particular sentence (or performing some other sign, e.g. a gesture) in a particular context. Speaker meaning requires the computation of *particularized conversational implicatures* (PCIs) to arrive at *what is (really) meant / implicated* (Grice, 1975). It is here where addressees must compute that saying "*The cat's bowl is empty*," might on this particular occasion be a request to feed her, a reproach, an attempt to share amazement at how fast can she eat, or just simple sharing of information; or that commenting on the dinner being formal also means disagreeing with the husband's choice of a tie. Speaker meaning is clear for the speaker, and must be recovered by the addressee on the basis of context. Context involves the speaker's referential intention, stance, and mutually known information, i.e. common ground (Stalnaker, 2002).

Typical listeners, who are not experts on pragmatic theories, generally think about "meaning" at this third level (Nicolle & Clark, 1999). If the husband reacted to the utterance meaning of the wife's remark and simply acknowledged that it was going to be a formal dinner, without doing anything about his tie, most people would conclude that he did not really understand her.

How do addressees come to understand speaker meaning if it is not "in" the actual words? According to Herbert Paul Grice, addressees routinely presume that speakers are rational and cooperative, and construct their utterances in a way that optimally expresses whatever they want to get across. Grice formulated a general principle of cooperation, according to which interlocutors should make their "*conversational contribution such as is required, at the stage at which it occurs, by the accepted direction or purpose of the talk exchange in which [they] are engaged*" (Grice, 1975, p.45). Addressees expect speakers to be relevant and truthful, and when speakers deviate from these expectations, addressees assume that their deviation has a purpose. When the husband hears a seemingly irrelevant comment from his wife instead of a straightforward "*yes*" or "*no*", he starts looking for possible reasons for this deliberate violating, or flouting, of tacit conversational rules. According to Grice, this is the point where the husband starts an inferential process to get at what his wife *really meant* with her remark.

Even though Grice's theory was not a psychological one, his explanation contains references to mental states (intentions) and cognitive processes (inferences). In interpreting speaker meaning, addressees start from the utterance itself, but ultimately try to *infer* the speaker's *intention*.

The idea that communication is not about decoding the meanings of words, but about inferences about the speaker's intentions was further elaborated by Sperber and Wilson (Sperber & Wilson, 1995; Wilson & Sperber, 2004). According to their Relevance theory, there are two types of intentions involved when people communicate. First of all, the speaker has an *informative* intention – an intention to provide the recipient with a new belief or to update an existing one (*the wife believes that the tie is inappropriate for the occasion*); and secondly, she also has a communicative intention – to make her informative intention known to her husband. The addressee's job is to recognize that the speaker's signal is communicative, and interpret the informative intention delivered together with it. This process is vastly different from a straightforward "receiving" and "decoding". There is a large number of possible informative intentions associated with each communicative signal, and the addressee has to narrow them down to the one that seems the most fitting in the given circumstances. In our model situation, the husband needs an opinion on his tie and even though the information is not explicitly present in his wife's answer, he can eventually narrow down the possible interpretations of his wife's remarks to the one that he is after – whether she thinks that the tie is a good idea. Pragmatic theories thus make a set of specific predictions as to what psychological processes are necessary to interpret a conversational remark, especially one whose speaker meaning is different from its utterance meaning. Apart from purely linguistic processing, comprehension would require processing related to interpreting mental states, such as beliefs and intentions.

Words are not the only vehicle for sharing our thoughts. The conversation between the couple can also take place in a different way. Imagine that while the husband asks for his wife's opinion, she is busy speaking on the phone. Instead of replying verbally, she might just point her finger at the extravagant tie with a disapproving look. Even though it is hard to talk about sentence meaning in this case – as there are no words and no syntactic structure – we can still partition her communicative act into several levels. She is singling out a specific part of reality (the tie), and she also wants to convey a specific message to her husband about it *("it is too informal for this occasion, change it")*. Also, she is extending her index finger *for* her husband, in reaction to his question, and not just randomly, or to inspect the color of her nail polish. Just like verbal utterances with sentence, utterance and speaker meaning, communicative gestures can also be "dissected" into conveying several levels of intention (Tomasello, 2008). Inferring what the gesture singles out in the world, i.e. the speaker's *referential intention* (akin to utterance meaning), is often not enough for successful communication. The key part is the speaker's *social intention* (or the speaker meaning of gesture); that is, what she wants to achieve by making that particular gesture in this particular situation – here, to motivate the husband to make another trip to the wardrobe.

Indirectness and directness explored here refer to the *correspondence between speaker and utterance meaning*; more precisely, between the speaker meaning that is superficially suggested by the utterance meaning and the one really meant by the speaker.

For the purposes of this thesis, I will define indirectness in the following way. In *direct* utterances, speaker meaning is expressed transparently/overtly, and therefore there is only a relatively small gap to be bridged between what was said and what was really meant. In *indirectness*, speaker meaning is not expressed transparently/overtly, and the addressee has to bridge a large gap between what is said and what is really meant by the speaker. For example, a declarative like "*It's raining,*" can involve an extra implicit declarative "*I don't want to go outside*". A declarative "*It's hard to give a good presentation*," implicates criticism of the speaker's performance; and, "*The cat's bowl is empty*", depending on the context, implicates another declarative ("*She eats a lot*") or a request ("*Feed her*").

3. Why do people speak indirectly?

While language gives us the means to express our ideas efficiently, the way we speak in everyday conversations may seem far from maximally efficient (Holtgraves, 2002). This has to do with the fact that communication is also a form of social action. Therefore, both *what* people say, as well as *how* they say it, has interpersonal implications: the delivery of the speaker's message is often tweaked to comply with a highly complex set of unwritten social rules. One of the most influential theories of linguistic politeness (Brown & Levinson, 1987) anchors language use in social action through the concept of *face* (Goffman, 1967). In interactions, partners typically work together to maintain a positive social image, or face, which includes respecting the addressee's desire for autonomy (negative face) and connection (positive face). Socially threatening conversational acts, such as requests or negative evaluations, are often conveyed with respect to such face concerns.

This interpersonal aspect of communication is one of the prime reasons why speakers make use of varying levels of directness. Typically, speakers avoid being direct when conveying messages that can be threatening for the addressee, or when they potentially limit his/her independence, such as in requesting ("*Would you mind passing the salt?*" instead of "*Give me that!*"). Blunt, short, direct answers are considered rude on most occasions, excluding life-threatening situations where fast action is crucial. The rules governing how and when directness is appropriate are determined by a complex, culture-specific set of expectations (e.g. status differences, interpersonal closeness, etc.). These expectations are normative, i.e. they are a learned set of implicit rules, much like Grice's cooperative principle, and conversational partners are expected to follow them.

In indirect utterances, there is a discrepancy between what the utterance meaning superficially suggest (what is said), and what the speaker really means. In socially sensitive situations, this discrepancy gives the speaker the opportunity to deny the offending/negative meaning (e.g. a threat – "*You've got a nice car here. It would be a shame if something happened to it*,"), since it is not part of what the addressee actually hears (Lee & Pinker, 2010; Pinker et al., 2008). In contrast, in direct communicative acts, the utterance meaning matches speaker meaning and is therefore much harder to deny (Holtgraves, 1986). Interpreting indirect replies places the burden of the inferential work on the addressee.

In addition to maintaining politeness in conversations or achieving plausible deniability, indirectness is also used in more mundane situations. Sometimes, the speaker wants to give extra information than specifically requested by her conversational partner, and thus does not answer a question with a simple yes/no, but rather with reasons for a "*yes*" or a "*no*". Such "information economy" (Walker, Drew, & Local, 2011) makes the job of the speaker slightly more difficult – she has to infer the yes/no herself – but also gives her extra valuable information. For example, when a person is asked whether she wants to join her friend for a run, she can reply by pointing out that her shoe is damaged, whereby she also provides the friend with a *reason* why she cannot join.

4. How do people interpret indirect meaning? Behavioral studies.

Understanding what people mean is almost never possible without considering the context in which the communicative act takes place. Ironically, some of the most decontextualized situations of language use can be found in psycholinguistic studies, where the need for experimental control urges researchers to protect their stimuli from all kinds of unwanted influences. Therefore, participants often find themselves in situations far removed from how they normally process language: they stare at rapidly presented words or sentences on the computer screen, or listen to the voice of an unknown speaker, without any relevance to them, or any consequences for their real life (for criticism of such an approach, see e.g. Van Berkum, 2009, 2012; Hoeks & Brouwer, 2014).`

Largely as a result of the methodological difficulties involved in studying more complex linguistic utterances, psycholinguistics has often ignored meaning above the utterance level. An exception is a relatively large body of studies focused on figurative language, such as metaphors ("*Lawyers are sharks*"), idioms ("*Kick the bucket*") or proverbs ("*Birds of feather flock together*"), and on conventional indirect requests ("*Can you pass the salt?*") (Holtgraves, 2002). All such expressions involve a gap between the conventional meaning of words or entire constructions, and their speaker meaning. In this way, they are indirect. However, there are important differences to the kinds of indirect replies that I focus on in this thesis, based on the way context contributes to their interpretation.

Indirect replies explored here are *particularized conversational implicatures* (PCIs) whose speaker meaning crucially depends on the context in which they are used – the message of "*It's hard to give a good presentation*," is very different when it comes after "*Did you like my presentation?*", then after "*Why do you prefer posters to presentations?*". In contrast, *generalized conversational implicatures* (GCIs), where most figurative meanings fall, can be interpreted in relatively context-free environments. Metaphors are a prime example of utterances that give rise to GCIs, as their meaning is derived from the semantic relations between the actual word meanings in the utterance, and the concepts in semantic memory that these words refer to, and thus independent of prior conversational context (Holtgraves, 2002). For example, "*Lawyers are sharks*", can be interpreted by finding a defining property of sharks – that they are aggressive – and mapping it to the characteristics typically valued in lawyers. In contrast, a PCI is activated only in appropriate contexts; thus, asking about a choice between a presentation and a poster does not activate the evaluative, negative aspect of "*It's hard to give a good presentation*" that arises after hearing such reply when specifically asking for feedback.

The reason why I am introducing this distinction here is that the overwhelming majority of studies on indirect meaning comprehension in fact zoomed in on *generalized* conversational implicatures. While they share the "utterance/speaker meaning" distinction with indirect replies, they are qualitatively different precisely in their lack of contextual dependence: to interpret a metaphor or an idiom, the addressee does not have to refer to the prior conversational context or consider the speaker's social motives. The other major difference between figurative meanings and indirect replies is that in experiments, figurative meanings are typically not used in a truly communicative way, to express a speaker's intention; usually, there is no speaker and no intention, with focus being solely on bridging the gap between the utterance meaning and the figurative (speaker) meaning of the expression.

One exception to the field's focus on figurative language in exploring speaker meaning interpretation is a small set of behavioral studies on PCIs by T. Holtgraves, who contrasted interpretation of direct and indirect replies. Holtgraves found that for indirect replies used as excuses and topic changes, politeness contexts activated their indirect readings. For most types of indirectness, this was an inferential, time-consuming process, during which their direct reading was also activated (Holtgraves, 1999).

This pattern of results – evidence about both direct and indirect reading of an indirect reply being processed – suggests that for PCIs, a Gricean-like analysis is taking place. The reply's utterance meaning is activated and when it turns out that it does not fit with the expectations created by the preceding discourse and social context, an inference is made to arrive at the implicated meaning. At the same time, activation of the utterance meaning does not automatically imply that the listener needs to first process it in its *entirety*. The addressee probably needs to encounter the *first* evidence that the utterance is not "going" in the expected direction (i.e. there is no clear relevance of the reply to the question, e.g. with a topic change *"How were your grades this semester?/It snowed very hard yesterday night,"*) to start the inferencing process to get at its implicated meaning. Holtgraves (1998) also showed that the social context in which indirect replies are used is crucial for their interpretation: when the preceding dialogue signaled that politeness expectations were at stake (i.e. indirect replies were meant as means of saving the addressee's face), participants were more likely to interpret irrelevant replies as conveying non-literal and negative information. If the face-saving reason for indirectness was not present (i.e. indirect reply was positive and therefore there was no clear interpersonal motive for indirectness), the interpretation process was more difficult, and took longer.

While these studies are clearly interesting, they cannot address the *identity* of the inferences necessary for indirect meaning interpretation. In the next section, we will turn to considering brain imaging research on non-literal meaning for more insights into this puzzle.

Box 1: Functional Magnetic Resonance Imaging

In three studies presented in this thesis, we look at "the brain in action" with the help of a neuroimaging technique called functional Magnetic Resonance Imaging (fMRI). An MRI scanner can produce both detailed images of the brain's anatomy, as well as a measure and visualization of brain activity related to cognitive processes taking place in different regions. Specifically, fMRI measures the Blood Oxygenation Level Dependent (BOLD) response of the brain. When neural tissue becomes active, brain cells in that region consume oxygen and that, in turn, is replenished via increased blood flow and blood volume. When the amount of blood in a region changes, there is also change in hemoglobin levels; due to changes in magnetic properties of oxygenized tissue, these changes can be measures as signal.

There are several practical considerations surrounding the use of fMRI for studying cognitive processes, including utterance comprehension. The BOLD response is slow to start, and it can take about 7 seconds to peak, even though the cognitive processes we want to measure typically unfold at a millisecond scale. Therefore, we use fMRI mainly to localize which brain regions are active in a particular task, and not to track how *fast* people process indirect meaning.

During language comprehension, there are a lot of different processes taking place at any given moment in the human brain, and most of them are not specific to indirectness. To isolate only the relevant ones, we use "subtraction logic": we not only measure the brain's response to indirectness, but also to a minimally different "baseline", i.e. carefully matched direct utterances. In this way, we can create statistical maps of the brain where we see activations for indirect processing, "cleaned" from processes implicated in interpreting direct utterances. This means that in the resulting subtraction images, we can see both regions that are implicated in both types of utterances but are *more* active for indirectness, as well as potential regions which are *only* active for indirectness. At the same time, we subtract out the processing of features of the utterances that are below the pragmatic level of meaning, and therefore not specific to pragmatic inferences. This is ensured by always making the direct and indirect utterances in our materials identical at the word- and sentence level.

Another practical issue inherent in fMRI research is that relative to everything else that is going on in the brain and in the body at any given moment, the signal of interest is very small. Due to this unfavorable signal-to-noise ratio, researchers must construct a large number of items and show each participant 25 to 30 of them per condition, in order to average over the whole set. This is also the reason why studies looking at the pragmatic level of processing tend to be time-consuming to prepare and carry out.

5. How do people interpret indirect meaning? Neuroimaging studies.

The core language network includes left-lateralized fronto-temporal areas in the perisylvian cortex, including Broca's and Wernicke's area (e.g. Hagoort, 2013; Hagoort & Poeppel, 2013). These are the regions responsible for the production and comprehension of propositional (sentence) meaning. But how does the brain deal with heavily context-dependent speaker meaning? There are two main venues in the neuroimaging literature into language comprehension above the utterance-level meaning. One deals with mostly figurative meanings, the other with comprehension of discourse, i.e. connected texts.

5.1 Neuroimaging studies of figurative expressions

Similarly as in the behavioral tradition, the most frequently studied types of indirect meanings have been figurative expressions, such as metaphors , idioms and proverbs; in addition, there are a number of studies on irony (*Wonderful weather for a picnic!* when it is actually raining). Based on two meta-analytic studies (Bohrn et al., 2012; Rapp et al., 2012), several points can be made about the different brain regions involved in their processing in comparison to literal[2] utterances.

First of all, there is, in general, greater *relative* involvement of the right hemisphere (RH) for nonliteral as opposed to literal expressions. However, in most studies involving metaphors and idioms (GCIs), the left hemisphere (LH) still dominates. RH involvement seems to correlate with non-salience: the more an expression is novel or non-salient, the larger the right hemispheric involvement. To explain RH activation, Beeman (Beeman, 1998; Jung-Beeman, 2005) proposed a *coarse semantic coding theory*, according to which the two hemispheres, given their anatomical differences, are more likely to code different semantic relationships. According to Beeman, the LH codes for finer, closer semantic relationships, and the RH, which is able to activate more diffuse semantic fields, codes for more distant, coarser semantic relationships. This might prove useful in natural language use where concepts are often distantly related, such as in figurative language, or when the reply is seemingly irrelevant to the answer, such as is the case with indirect replies.

Secondly, not all figurative meanings are "created equal": there are differences in processing based on the identity of a figure of speech, as well as on its conventionality. While conventional metaphors ("*Lawyers are sharks"*) tend to activate the left frontotemporal regions, processing related to novel/nonsalient metaphors (*his smile was a ray of sunshine*) is more right-lateralized, activating also the right inferior frontal gyrus (IFG) and anterior cingulate cortex. In comparison to metaphors, the activation of the left IFG for idioms and proverbs is more dorsolateral, and they also activate the right IFG region. Irony and sarcasm, which are often studies in this tradition although they are not strictly figurative, seem to result in the most distinct pattern of activation, including the medial prefrontal cortex, the anterior cingulate cortex (ACC) and the right superior temporal gyrus.
Third, apart from the core fronto-temporal network implicated in language processing, there are relatively few "extra areas" in comparison to literal language processing, suggesting that non-literal processing uses largely the same networks. For most figurative meanings, additional activations were seen in regions implied in choosing the appropriate meaning

[2] I will use the expressions "literal" in this section to refer to the non-figurative (utterance-level) readings of utterances, as is the usual term in this research tradition.

among many, or integrating this meaning into the expression – the left anterior and posterior middle temporal gyrus (Bohrn et al., 2012).

In what way can these results speak to where to expect differences in brain activations for indirect versus direct utterances? On the one hand, as I mentioned in the previous section, figurative expressions are similar to indirect expressions because in both cases, their utterance meaning and speaker meaning do not match. On the other hand, the way listeners use context – and what kind of context they refer to - in bridging the explicit-implicit gap is likely quite different, since most of these figurative expressions are GCIs, and often also conventional (e.g. they have a set meaning and are learned by heart, not invented by the speaker each time she uses such expression). Indirect replies are PCIs and therefore entirely context-dependent. This implies that while PCIs would be more taxing for right-lateralized regions, GCIs for brain regions involved in conceptual access. Most importantly, only irony and sarcasm make reference to the speaker's intentions, since they convey the speaker's attitude towards something or someone. This qualitative difference in processing is also evident in the activation patterns, with irony and sarcasm activating regions typically involved in social cognition.

5.2 Neuroimaging studies on contextual integration

Another way of accounting for the reliance on contextual information in the interpretation of utterances comes from a metaanalysis on discourse comprehension by Evelyn Ferstl and colleagues (Ferstl, Neumann, Bogler, & von Cramon, 2008) who aggregated studies on comprehension of multi-sentence units. In order to understand connected text, listeners or readers must make good use of contextual sources: they must connect adjacent sentences, trace referents, fill in implicit information using knowledge about the world. This activates regions beyond the core language areas, with different contrasts yielding different patterns. Comparing connected, comprehensible texts to a control condition lacking any plot or story (e.g. word lists) yielded bilateral activation in the anterior temporal lobes, as well as in the left middle temporal gyrus, left inferior frontal cortex and left medial prefrontal cortex. In another contrast, the authors analyzed a small number of "special" activations, for example interpreting the moral message of the story. Again, bilateral anterior temporal lobes and the left inferior frontal gyrus were activated, as well as the right lateral prefrontal cortex and the temporo-parietal junction. These analyses, although fairly general in scope of the involved studies, show that as soon as language comprehension goes beyond the utterance-level meaning -- here, in terms of multi-sentence units -- the core language network is not sufficient. However, there are still important differences to be expected between the processes involved in interpreting multi-sentence utterances and speaker meaning. Most importantly, the vast majority of studies in Ferstl's metaanalysis did not require participants to reason about the speaker's intentions, as there were no speakers or conversations.

6. What is missing and how can we address it?

Coming back to our example of a communicative exchange between the husband and his wife, and comparing it to studies on figurative meanings and multi-sentence utterances, there seems to be something missing. The fact that the husband was able to interpret his wife's seemingly irrelevant remark as an indirect suggestion to wear a different tie is – in Grice's and Relevance theory's theoretical analysis – crucially supported by his ability to get at his wife's intentions. Studies on figurative meaning and multi-sentence discourse typically do not look at this level of processing. In fact, as I already pointed out, there is typically no speaker and no message in the majority of neurocognitive studies on language comprehension. At the same time, studies on intention processing do not focus on language comprehension. Those that use verbal stimuli treat language just as a vehicle for delivering the message, not as the "puzzle" in itself (i.e. studies on mentalizing, next section).

In this sense, studying the neural correlates of indirect utterance comprehension seems like a good opportunity to explore the inferences necessary for interpreting speaker meaning. There is a large gap between their direct and indirect meaning which the addressee has no other way to bridge than to use all the available contextual information, including tacit knowledge about politeness strategies in interactions.

Box 2: The experimental paradigm

In all four studies reported in this thesis, I use the same experimental paradigm. In order to isolate only processing related to indirectness, I chose a closely-matched baseline, direct replies. The design relies on dialogs, where the preceding question naturally renders the target utterance direct or indirect – without changing the choice of words or sentence structure. So, "*It's hard to give a good presentation*," becomes either direct, informative answer to "*How difficult is it to give a good presentation?*", or an indirect criticism to a friend's not-so-optimal performance after "*Did you like my presentation?*".

In order to have as much control as possible over the contextual information listeners have at hand to guide their interpretation of the indirect replies, each dialog is preceded by a short spoken or written context story. Here, the most relevant pieces of information about the protagonists and the situation are introduced. There is also always a task to keep the listeners interested, and to check whether they understood the indirect message. In Chapter 2, they have to answer probe questions on filler items, in Chapter 3, they evaluate speakers who gave indirect replies based on information gleaned from these replies. Finally, in Chapters 4 and 5, participants have to decide whether a sentence was an appropriate or logical continuation of the dialog.

7. Mentalizing and the brain

Another domain that is relevant to speaker meaning comprehension falls outside the scope of language research. Successfully navigating the social world involves understanding that people have beliefs, wishes, desires or emotions, and acting on this understanding when explaining and predicting their behavior. The ability to make inferences about the mental states of others, as well as one's own, is called mentalizing or theory of mind (ToM), and has been extensively studied in both psychology (e.g. Frith & Frith, 2010; Saxe, 2006) and cognitive neuroscience (e.g. Gallagher & Frith, 2003; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). The most important developmental milestone of ToM is the ability to understand that people can have beliefs that are false, and still act on them (the false-belief task, Wimmer & Perner, 1983).

Even though pragmatic theories predict that inferring the speaker's communicative intention should involve some form of mentalizing (Wilson & Sperber, 1986), not many actual neuroimaging studies have confirmed it. There might be several reasons for this: for example, the overall focus on GCIs in research on non-literal language processing, which can be interpreted with limited access to conversational context or speaker's intentions. In addition, studies are set up in a way that there are no real communicative situations, and no speaker whose intention one can infer (as Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012 argue for irony research). Therefore, despite clear theoretical predictions, the actual role of mentalizing in inferring the speaker's intention is still unclear.

Investigating pragmatic inferences with fMRI should give us a chance to see whether interpretation of indirect replies activates regions associated with ToM processing. However, there are also caveats: while we know that ToM is subserved by a network of brain regions, the functional role of each of them is unclear at present. There is also substantial variability in activations in the large number of published studies. One of the reasons is that researchers use a great variety of tasks and stimulus formats, and this leads to different patterns of brain activity (Mar, 2011; Molenberghs, Johnson, Henry, & Mattingley, 2016; Schaafsma, Pfaff, Spunt, & Adolphs, 2015; Schurz, Aichhorn, Martin, & Perner, 2013; Schurz et al., 2014). For example, Schurz at al. (2014) identified 6 different activation profiles based in task patterns, such as for false beliefs vs false physical representations (photographs), strategic games, estimating personal traits based on behavior, judging mental states from photographs of eyes, or rational actions.
Based on various meta-analytic techniques, we now have fairly strong evidence that despite this variability, there is a set core areas activated whenever people reason about the mental states of others, regardless of the task or stimuli used (Schurz et al., 2014, 2015; Molenberghs et al., 2015). These core areas include the medial prefrontal cortex (mPFC) and temporo-parietal junction (TPJ).

8. Summary and preview

In sum, this thesis will be an investigation of how the brain deals with interpreting particularized conversational implicatures triggered by indirect replies. In Chapter 2, I start out

by contrasting direct utterances with indirect utterances, both embedded in dialogs that the participants are overhearing. While politeness is traditionally viewed as one of the main reasons why speakers sacrifice efficiency in communication, there are also other types of motivation to speak indirectly. To cast the net wider, I therefore contrast two classes of indirect replies: face-saving replies, and more neutral indirect replies prompted by information economy. One of the issues investigated in this study is also whether interpreting indirectness requires mentalizing inferences, or whether simple simulation or mirroring is sufficient.
In Chapter 3, I will again focus on the interpersonal aspect of indirectness, and ask how the brain response changes when the participant is no longer an overhearer, but an active addressee of a face-saving indirect reply. This issue touches on frequent methodological criticism of psycholinguistic and neurocognitive studies – namely, that participants are virtually always just overhearers of language that is not personally relevant to them in any way (Hoeks & Brouwer, 2014; Van Berkum, 2009; 2012). In the study reported in this chapter, I set up a mock job interview, in which participants have to ask questions to a number of "job candidates" and judge them on their suitability for several positions. Since a job interview is quite naturally a situation where people want to present themselves in a better light, some of the candidates' answers are indirect, because they are trying to cover up for lack of ability or to make themselves more desirable. I contrast two situations: indirect replies where the candidate addresses the participant directly, and identical indirect replies which are addressed to a third person, so that the participant is just an overhearer. The crucial difference is that in the "addressee" case, the speaker is trying to "play" the participant him- or herself. In contrast, in the overhearer case, the candidate's impression management is directed at someone else. My prediction is that placing the participants in the role of active addressees of face-saving indirect replies will not fundamentally change the "inferential architecture" behind indirectness, but might influence how emotionally engaged the addressees are.

In studies reported in Chapters 4 and 5, I take a step back from the linguistic code and look at the "bigger picture". Here, the stimuli are indirect and direct replies delivered by verbal utterances or pointing and showing gestures. By comparing matched indirect replies in two modalities, I aim to determine what is the time-course of interpretation of gestural and verbal indirectness (Chapter 4, reaction times) and which are the "core" indirectness regions, independent of the modality in which they are delivered to the addressee (Chapter 5, fMRI). I hypothesize that the brain response to the speaker's intention behind indirect communicative signals should be identical regardless of the modality that they are delivered in. If anything, verbal and gestural indirectness might elicit a differential response in regions involved in inferring the speaker's referential, not social, intention. Methodologically, these two studies take a simpler approach, as I use simple cartoons depicting various real-life situations as stimulus items, instead of auditory recordings, and "informative" indirectness instead of face-saving one.

In the studies reported in this thesis, I place great emphasis on *replication*, which is currently an object of considerable discussion in psychology and cognitive neuroscience (Open Science Collaboration, 2015). All three fMRI studies test the same indirectness contrast (indirect replies > direct replies). I employ the same paradigm, but different stimuli and participants; there are also differences in the modality of presentation, as indirect replies are delivered via auditory recordings in Chapters 2 and 3, and as written texts in Chapters 4 and 5.

References

Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In M. Beeman & C. Chiarrello (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience.* (pp. 255–284).

Brown, P., & Levinson, S. C. (1987). *Politeness: some universals in language usage*. Cambridge University Press.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*(3), 567–598. https://doi.org/10.1080/01638539709545003

Ferstl, E., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A metaanalysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581–593.

Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1537), 165–76. https://doi.org/10.1098/rstb.2009.0160

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Sciences*, *7*(2), 77–83.

Garrett, M., & Harnish, R. M. (2007). Experimental pragmatics: Testing for implicitures. *Pragmatics & Cognition*, *15*(1), 65–90. https://doi.org/10.1075/pc.15.1.07gar

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday Anchor.

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

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax And Semantics* (Vol. 3, pp. 41–58). Academic Press.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*.

Hagoort, P., & Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, *37*(1), 347–362. https://doi.org/10.1146/annurev-neuro-071013-013847

Hagoort, P., & Poeppel, D. (2013). The Infrastructure of the Language-Ready Brain. *Language, Music, and the Brain*.

Hoeks, J. C., & Brouwer, H. (2014). Electrophysiological Research on Conversation and Discourse. In T. M. H. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology* (pp. 365–386). New York: Oxford University Press.

Holtgraves, T. (1986). Language Structure in Social Interaction. Perceptions of Direct and Indirect Speech Acts and Interactants Who Use Them. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.51.2.305

Holtgraves, T. (1998). Interpreting indirect replies. *Cognitive Psychology*, *37*(1), 1–27.

Holtgraves, T. (1999). Comprehending indirect replies: When and how are their conveyed meanings activated? *Journal of Memory and Language*, *41*(4), 519–540. https://doi.org/https://doi.org/10.1006/jmla.1999.2657

Holtgraves, T. M. (2002). *Language as social action: Social psychology and language use.* Mahwah, NJ.: Erlbaum.

Jackendoff, R. (2007). *Language, consciousness, culture: Essays on mental structure.* MIT Press.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–8. https://doi.org/10.1016/j.tics.2005.09.009

Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: the theory of the strategic speaker. *Psychological Review*, *117*(3), 785–807.

Mar, R. A. (2011). The Neural Bases of Social Cognition and Story Comprehension. *Annual Review of Psychology*, *62*(1), 103–134. https://doi.org/10.1146/annurev-psych-120709-145406

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Nicolle, S., & Clark, B. (1999). Experimental pragmatics and what is said: A response to Gibbs and Moise. *Cognition*, *69*(3), 337–354. https://doi.org/10.1016/S0010-0277(98)00070-5

Noveck, I. A., & Reboul, A. (2008). Experimental Pragmatics: a Gricean turn in the study of language. *Trends in Cognitive Sciences*, *12*(11), 425–431. https://doi.org/10.1016/J.TICS.2008.07.009

Open Science Collaboration, O. S. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(3), 833–838. https://doi.org//10.1073/pnas.0707192105

Reddy, M. (1979). The conduit metaphor. *Metaphor and Thought*, *2*, 285–324.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–9. https://doi.org/10.1016/j.conb.2006.03.001

Shannon, C. E. J. (1948). A mathematical theory of communication. *Bell Syst. Tech.*, *27*, 379–423–656.

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, *19*(2), 65–72. https://doi.org/10.1016/J.TICS.2014.11.007

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience*, *7*, 712.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral*

Reviews, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Sperber, D., & Wilson, D. (1995). Relevance: communication and cognition. *Behavioral and Brain Sciences*, *2nd*(4), 697–710. https://doi.org/10.1017/S0140525X00055345

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39. https://doi.org/https://doi.org/10.1016/j.neuroimage.2012.06.046

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, *25*(5–6), 701–721.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Van Berkum, J. J. A. (2009). The neuropragmatics of “simple” utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave Macmillan.

Van Berkum, J. J. A. (2012). *The Cambridge Handbook of Psycholinguistics*. (M. Spivey, K. McRae, & M. Joanisse, Eds.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139029377

Walker, T., Drew, P., & Local, J. (2011). Responding indirectly. *Journal of Pragmatics*, *43*(9), 2434–2451. https://doi.org/10.1016/j.pragma.2011.02.012

Wilson, D., & Sperber, D. (1986). Pragmatics and modularity. *In: Farley, A and Farley, T and McCullough, K, (Eds.) (Proceedings) Chicago Linguistic Society 22. (Pp. Pp. 67-84). Chicago Linguistic Society: Chicago, IL. (1986)* .

Wilson, D., & Sperber, D. (2004). Relevance Theory. In G. Horn, L.R. & Ward (Ed.), *The Handbook of Pragmatics* (pp. 607–632). Oxford: Blackwell.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*. https://doi.org/10.1016/0010-0277(83)90004-5

CHAPTER 2: NEURAL CORRELATES OF INDIRECT REPLIES

Beyond the language given: The neural correlates of inferring speaker meaning

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

[1] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands [2] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands [3] Utrecht University, UiL-OTS and Department of Dutch, Trans 10, 3512 JK Utrecht, The Netherlands[4] Institute of Experimental Psychology, SAS, Dubravska cesta 9, 813 64 Bratislava, Slovakia

Abstract

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. One of the big puzzles in language science is how listeners work out what speakers really mean, a skill absolutely central to communication. However, most neuroimaging studies of language comprehension have focused on the simpler, context-independent process of understanding direct utterances. We used fMRI as people listened to indirect replies in spoken dialogue. Relative to direct control utterances, they engaged dorsomedial prefrontal cortex, right temporo-parietal junction and insula, as well as bilateral inferior frontal gyrus and right medial temporal gyrus. This implies that listeners take the speaker's perspective on both cognitive (theory-of-mind) and affective (empathy-like) levels. Our results indicate that currently popular "simulationist" accounts of language comprehension fail to explain how listeners understand the speaker's intended message.

Based on:
Jana Bašnáková, Kirsten Weber, Karl Magnus Petersson, Jos van Berkum, & Peter Hagoort (2014). Beyond the language given: The neural correlates of inferring speaker meaning. Cerebral Cortex, Vol. 24, Iss. 10, 1 October 2014, Pages 2572–2578.

1. 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 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 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 towards 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 - two 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 dialogue in which the final and critical utterance, e.g., "*It is hard to give a good presentation*", had different meanings depending on the dialogue 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 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. Brown and Levinson, 1987; Goffman, 1967; 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 three 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; Wilson and Sperber, 2004; Levinson, 2000). 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 (ATL) (Amodio and Frith, 2006; Mitchell et al., 2006; Saxe, 2006; Frith and Frith, 2003; Saxe and Kanwisher, 2003). Since the communicated meaning of indirect replies also depends on non-mentalistic 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 two 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 (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 which 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. Buccino et al., 2001; Iacoboni, 1999; Iacoboni et al., 2001; Rizzolatti et al., 1996).

In the domain of understanding language, the putative mirror neuron system has been argued to be implicated in, for example, action word comprehension (e.g. Hauk et al., 2004), 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 two people while their 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.

2. Results

The crucial comparison for the issue at stake is the one between same sentences in their role as indirect replies 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 inferior frontal gyrus (pars orbitalis and pars triangularis) bilaterally, extending into the insula in the left hemisphere (see Figure 1; for exact coordinates, see Table 1). In addition, the right temporo-parietal junction (TPJ) and the right middle temporal gyrus showed an increased activation. In order to disentangle the relative contribution of the two 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 activation included the left anterior cingulate (ACC), and an additional activation cluster was present in the right superior temporal gyrus (STG) (see SI, Figure S1.B). When we excluded the face-saving indirect replies, still a significant subset of the regions from the pooled comparison remained active: the right temporo-parietal junction, 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 (Figure S1.A).

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 superior temporal gyrus and the anterior cingulate cortex. Once again, there was a cluster spanning the right inferior frontal gyrus and anterior insula, although this time it was only marginally significant, at $p < 0.10$ (Figure S1.C). Results and figures from the conjunction analysis of the two indirect effects are presented in SI (S1.D, S2).

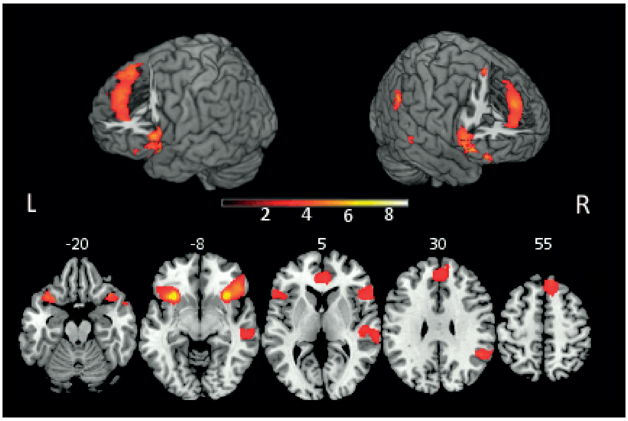


Figure 1: Activations for the main effect of indirectness. Significant effects are displayed on cortical renderings and on axial slices (z coordinate levels in millimeters). (See Figure S1 for additional effects.)

Anatomical region	Coordinates of local maxima			BA	Cluster size	p-value (cluster-level FEW corrected)
	x	*y*	*z*			
Pooled indirect replies > direct replies						
R supplementary motor area	14	24	58	8	1614	<.001
L middle frontal medial gyrus	-4	42	28	32		
R middle temporal gyrus	8	36	50	6		
R inferior frontal gyrus	34	22	-12	47/49	1592	<.001
	44	30	-4	47		
	58	22	4	47		
L anterior insula	-30	16	-14	15	1381	<.001
L inferior frontal gyrus	-58	22	10	47		
	-52	38	-6	47		
R temporo-parietal junction	48	-50	32	39/40	375	.003
R middle temporal gyrus	52	-38	-6	21	248	.022
	62	-30	-2	21		
Indirect informative replies > direct replies						
L temporal pole	-36	16	-20	38	472	.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	.011
R inferior frontal gyrus (pars orbitalis)	38	22	-12	47/49	189	.049
Indirect face-saving replies > direct replies						
R anterior cingulate cortex	6	44	16	24/32	2137	<.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	<.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	<.001

L inferior frontal gyrus (pars triangularis)	-58	24	8	45		
R superior temporal gyrus	58	-22	2	22	763	<.001
	64	-28	0	21/22		
R middle temporal gyrus	54	-28	-4	21		
R temporo-parietal junction	50	-50	32	39/40	247	.030
Indirect face-saving replies > indirect informative replies						
R superior temporal gyrus	56	-20	4	41/42/22	1011	<.001
	50	-26	8	41/42/22		
	56	-6	-8	22		
R anterior cingulate cortex	6	48	18	32	414	.002
R inferior frontal gyrus (pars orbitalis)	44	26	-14	47	166	.091
R anterior insula	32	20	-16	15		

Table 1. Activations for contrasts of interest thresholded at 0.001. Cluster P-values are corrected for multiple non-independent comparisons. All reported coordinates are in MNI space. (See Table S2 for conjunction analysis activations.)

3. Discussion

The main goal of this study was to identify brain regions involved in language comprehension at a level of representation typically overlooked in 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 which 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 (medial frontal cortex (MFC), right temporo-parietal junction (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 medial frontal cortex extending into the right anterior part of SMA, and in the right temporo-parietal junction, a pattern typical for tasks which 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, i.e. 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).

The MFC cortex is a large cortical region with a variety of roles characteristic of social cognition in general, beyond ToM processing (Saxe and Powell, 2006; Amodio and Frith, 2006). Based on a meta-analysis of task-related activations from this region, Amodio and Frith (Amodio and Frith, 2006) proposed a division of the MFC into three 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, right anterior MFC) or oneself (right posterior MFC). Interestingly, the involvement of these regions is not exclusive to theory-of-mind tasks, but is consistently observed in the narrative comprehension literature (e.g. Mar, 2011; Mason and Just, 2009). 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 two brain regions typically involved in social cognition suggests that listeners engage in non-linguistic 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 meanings of her words, but also what was her motivation and what goal she wanted to achieve when she uttered these words in 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 Saxe, 2005), 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 labour 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 two types of non-linguistic 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 SI (S4).

Also involved in recovering the meaning of indirect replies were parts of the perisylvian language network in the left inferior frontal gyrus (BA45 and 47) and their right-hemisphere homologues. 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 non-linguistic input (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.

3.1. 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 which 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 which 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 explaining 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 two 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 (Saxe, 2005, 2009; Decety, 2010; Picard and Strick, 2001). 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.

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 (theory-of-mind) and affective (empathy-like) levels.

4. Experimental Procedures

4.1. Participants

28 native speakers of Dutch participated in the experiment (5 male, 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.

4.2. Stimulus material

We created 90 critical utterances which were preceded by three different types of context, making up 270 experimental items in total. There were three 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 socio-emotional 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 (Holtgraves, 1999).

The structure of each item was as follows: a lead-in story set up the relevant context, introducing the two 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 three conditions, the critical utterance (reply) was always the same. Thus, we had two 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 S3).

In addition, there were 55 filler items. The purpose of the filler items was two-fold: firstly, 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 and Weenink, 2009) and two native Dutch speakers then jointly chose the best recording for each of the three conditions.

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.***

Table 2: Examples of the three different types of replies, preceded by their respective context stories, translated into English. The target utterance is always the final one (in ***bold italics***).

4.3. Procedure

The study consisted of two sessions lasting approximately 1.5 hours each, on two different days. One session comprised two experimental blocks, the other three 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 seconds. The lead-in story was then presented in stereo via headphones. After the last sentence of the story, the names of the two protagonists depicted in the story were displayed on the screen for another 2 seconds, one on the left and one 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 three stimulus lists.

We matched the three 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 (GCN, 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 pseudo-randomized, with no more than 2 items from the same condition appearing after each other.

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

4.4. 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.

4.5. fMRI data acquisition

Participants were scanned in an ascending fashion with a Siemens 3T 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 3.5x3.5x3

mm^3 and the field of view was 224 mm. Functional scans were acquired at TE=30 ms. Flip angle was 80 degrees. A whole-brain high resolution structural T1-weigthed MPRAGE sequence was performed to characterize participants' anatomy (TR = 2300 ms, TE = 3.03 ms, 192 slices with voxel size of 1 mm^3, FOV = 256).

4.6. fMRI data analysis

The 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 non-saturation effect from affecting the analysis. The functional EPI-BOLD images were then realigned and slice-time corrected. The resulting functional images were co-registered to the participants' anatomical volume based on the subject-mean functional image, normalized to MNI space and spatially smoothed using a 3D isotropic Gaussian smoothing kernel (FWHM =8 mm). A temporal high-pass filter was applied with a cycle cut-off at 128s.

In the first level linear model, we modeled the onsets and durations of the 3 types of the target utterances (Direct, Indirect neutral and Indirect informative), which were defined as the entire conversational turn, including the short pre-utterance 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 HRF 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 2nd-level random effects analysis we used the contrast-images of interest in a repeated measures ANOVA. The cluster size was used as the test statistic and only clusters significant at $P < 0.05$ corrected for multiple non-independent comparisons are reported.

References

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. Nature reviews. Neuroscience *7*, 268–277.

Berntson, G. G., Norman, G. J., Bechara, A., Bruss, J., Tranel, D., and Cacioppo, J. T. (2011). The insula and evaluative processes. Psychological Science *22*, 80–86.

Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. Trends in Cognitive Sciences *15*, 527–536.

Boersma, P., and Weenink, D. (2009). Praat: doing phonetics by computer. Computer program version 4322.

Brown, P., and Levinson, S. C. (1987). Politeness (Cambridge Univ Press).

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., and Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. European Journal of Neuroscience *13*, 400–404.

Carrington, S. J., and Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. Human Brain Mapping *30*, 2313–2335.

Dalgleish, T. (2004). The emotional brain. Nature reviews. Neuroscience *5*, 583–589.

Decety, J. (2010). To What Extent is the Experience of Empathy Mediated by Shared Neural Circuits? Emotion Review *2*, 204–207.

Fan, Y., Duncan, N. W., de Greck, M., and Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. Neuroscience and Biobehavioral Reviews *35*, 903–911.

Ferstl, E., Neumann, J., Bogler, C., and von Cramon, D. Y. (2008). The extended language network: A metaanalysis of neuroimaging studies on text comprehension. Human Brain Mapping *29*, 581–593.

Frith, U., and Frith, C. D. (2003). Development and neurophysiology of mentalizing. Philosophical Transactions of the Royal Society B Biological Sciences *358*, 459–473.

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

Grice, H. P. (1975). Logic and Conversation. In Syntax And Semantics, P. Cole and J. L. Morgan, eds. (Academic Press), pp. 41–58.

Grice, H. P. (1957). Meaning. The Philosophical Review, 377–388.

Hagoort, P., Baggio, G., and Willems, R. (2009). Semantic unification. In The cognitive neurosciences, M. S. Gazzaniga, ed. (Cambridge, MA: MIT Press), pp. 819–836.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. Neuron *41*, 301–307.

Holtgraves, T. (1999). Comprehending indirect replies: When and how are their conveyed meanings activated? Journal of Memory and Language *41*, 519–540.

Iacoboni, M. (1999). Cortical Mechanisms of Human Imitation. Science *286*, 2526–2528.

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M.-C., Mazziotta, J. C., and Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. Proceedings of the National Academy of Sciences of the United States of America *98*, 13995–13999.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., and Holcomb, P. J. (2006). Making sense of discourse: an fMRI study of causal inferencing across sentences. NeuroImage *33*, 343–361.

Levinson, S. C. (2000). Presumptive Meanings (Cambridge, MA: MIT Press).

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., and 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, R. A. (2011). The Neural Bases of Social Cognition and Story Comprehension. Annual Review of Psychology.

Mar, R. A., and Oatley, K. (2008). The Function of Fiction is the Abstraction and Simulation of Social Experience. Perspectives on Psychological Science *3*, 173–192.

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

Mason, R. A., and Just, M. A. (2011). Differentiable cortical networks for inferences concerning people's intentions versus physical causality. Human Brain Mapping *32*, 313–329.

Mason, R. A., and Just, M. A. (2009). The Role of the Theory-of-Mind Cortical Network in the Comprehension of Narratives. Language and Linguistics Compass *3*, 157–174.

Menenti, L., Petersson, K. M., Scheeringa, R., and Hagoort, P. (2009). When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. Journal of Cognitive Neuroscience *21*, 2358–2368.

Mitchell, J. P., Macrae, C. N., and Banaji, M. R. (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., and Strick, P. L. (2001). Imaging the premotor areas. Current Opinion in Neurobiology *11*, 663–672.

Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. Journal of Anatomy *197 Pt 3*, 335–359.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. Brain Research *3*, 131–141.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. Annual review of neuroscience *27*, 169–192.

Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. (Nature Publishing Group).

Saxe, R. (2005). Against simulation: the argument from error. Trends in Cognitive Sciences *9*, 174–179.

Saxe, R. (2009). The neural evidence for simulation is weaker than I think you think it is. Philosophical Studies *144*, 447–456.

Saxe, R. (2010). The right temporo-parietal junction: a specific brain region for thinking about thoughts. In Handbook of Theory of Mind, A. Leslie and T. German, eds. (Psychology Press).

Saxe, R. (2006). Uniquely human social cognition. Current Opinion in Neurobiology *16*, 235–239.

Saxe, R., Moran, J. M., Scholz, J., and Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. Social Cognitive and Affective Neuroscience *1*, 229–234.

Saxe, R., and 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., and Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. Psychological Science *17*, 692–699.

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

Singer, T., and Lamm, C. (2009). The social neuroscience of empathy. Annals of the New York Academy of Sciences *1156*, 81–96.

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

Wilson, D., and Sperber, D. (2004). Relevance Theory. In The Handbook of Pragmatics, G. Horn, L.R. & Ward, ed. (Oxford: Blackwell), pp. 607–632.

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

Acknowledgment

The first author was partly supported by grant agency VEGA (2/0204/09).

Supplementary materials

S5: Construction of the stimulus material

In order to achieve a direct or an indirect reading of the critical utterance, we applied the following rules when constructing the experimental items:

(i) The critical question in the direct condition was an open-ended, wh- question, followed by a direct reply. The two indirect conditions were preceded by a yes/no question, and upon hearing the critical utterance, the listener was forced to interpret the reply as either a 'yes' or a 'no'. Since most of the naturally occurring face-saving indirect replies are negative (refusals, criticism), in order to make the two indirect conditions comparable, we constructed *all* the indirect replies as implying a "no". In order to balance the design, the filler items mostly implied, or explicitly stated, a "yes" reply.

(ii)While the stories and dialogs in the direct and indirect informative conditions were as emotionally neutral as possible, we made sure that in the indirect face-saving condition, protagonists had always "something at stake". The utterances in the indirect face-saving condition had a clear social-emotional implication – they were meant as polite refusals, excuses, as criticism, etc. Thus, although both types of indirect replies needed to be processed for indirect meaning in order to provide a satisfactory interpretation in light of the preceding yes/no question, the *reason* for indirectness differed. In the indirect informative condition, the protagonist did not answer directly because he/she was simply providing more information, while in the indirect face-saving condition, the protagonist did not answer directly because he/she was trying to avoid a direct (negative) answer in order to save his own or his/her conversational partner's face.

The experimental items were constructed in several steps. First, each (written) item was subjected to feedback from two independent judges and changed in accordance with their suggestions. This step was repeated twice. After that, a written version of the 270 stimuli was given to 9 naïve participants, who were asked to choose the best paraphrase of the critical utterance' *intended* meaning in light of the preceding context (coded as a direct literal reply, indirect informative reply, indirect face-saving reply, neither). There were 3 stimulus lists, with each critical sentence appearing on each list only once. Therefore each item was rated three times. If an item was misclassified more than once, it was subsequently discussed by three judges and modified.

In the next step, the dialogs were recorded and the narrators were asked to report if the items sounded unnatural or inconsistent to them. Again, the items were changed accordingly.

The last check was when two native Dutch speakers listened to all the recorded material and chose the best instance of each dialog and critical utterance. At this stage again, problematic items were changed and re-recorded. The final set of stimuli was edited in Praat, adjusting the volume and cutting the dialog into conversational turns. Even though the individual lines of the dialogs were sometimes selected from several different recording trials, we never used a target utterance from a different condition than the one in which it was originally recorded. This was done in order to preserve subtle timing and prosodic characteristics of the utterances.

Although the lead-in stories and dialogs leading up to the critical utterances were not identical, we took great care to make them as similar as possible and matched them on a number of features: (i) length of the lead-in story and the dialog in words (direct = 94.2, indirect informative = 95.7, indirect face-saving = 95.5); (ii) the extent of direct semantic priming – i.e. how many content words from both the context and the critical question were repeated in the critical utterance; (iii) loudness of the critical utterance across conditions; (iv) semantic similarity of the lead-in story and dialog to the critical utterance. Semantic similarity was compared by means of latent semantic analysis (Landauer and Dumais 1997; Landauer et al. 1998) (available online at http://lsa.colorado.edu). Since there are currently no reasonably large semantic corpora available in Dutch, we translated the stimuli into English and carried out the analysis in English. A term-to-term comparison using tasaALL space, which corresponds to a reading level of a third-grade college student, yielded similar mean semantic similarity values (SSVs): direct = 0.769, indirect informative = 0.768, indirect face-saving = 0.763.

The critical utterances were three different tokens of the same type and although they were identical at the word and sentence level, there were minimal differences in the duration of each conversational turn and in intonation. We chose not to make these properties uniform or to use a single token of the utterance in all three conditions, since we wanted to mimic the properties of natural human communication.

S6: Processing of face-saving indirect replies

What additional processes work together in inferring speaker meaning when the primary motivation for indirectness is to not hurt listener's feelings? The most salient feature of face-saving indirect replies is their emotional connotation – all of them conveyed essentially negative information in situations when the listeners had "something at stake". Thus, we expected that they will lead to more activation in emotion-related areas, and this was indeed what we found: listening to excuses and polite refusals lead to the involvement of ACC and insula, which are associated with empathy, mentalizing and emotional processing (Bush et al. 2000; Decety and Jackson 2004; Van Overwalle 2009; Fan et al. 2011).

There were, however, brain areas which are associated with functions outside the social-emotional domain: the right superior temporal gyrus and the right IFG, related to discourse-level inferencing. Activation in rIFG increases with increasing difficulty of integrating incoming information into context (e.g. Tylén et al. 2009) which might result in a more effortful process of building the model of the on-going discourse (Menenti et al. 2009; Tesink et al. 2009). Increased signal in the rSTG is observed in studies where participants need to supply information which is not explicitly stated, or connect different sources of information during inference generation (Virtue et al. 2006), and together with the right middle temporal cortex, they constitute the coarse semantic processing network within the model of discourse comprehension by Mason and Just (Mason and Just 2006).

In the context of face-saving indirect replies, this extra layer of complexity which triggers additional inferencing and contextual integration might be the socially relevant information, such as knowledge of social norms, which is necessary to successfully interpret their implicit message. It is precisely the recognition of the social motivation for indirectness which helps the listener to arrive at the intended, indirect meaning of the reply: the listener can start her search for relevance upon recognition that the speaker is trying to convey his reply in a considerate manner.

Interestingly, all of the activations outside the core indirectness areas are right-lateralized – this is in line with studies on text comprehension suggesting that as language comprehension gets more complex, there is increasing involvement of right-hemisphere areas (Jung-Beeman 2005; Xu et al. 2005).

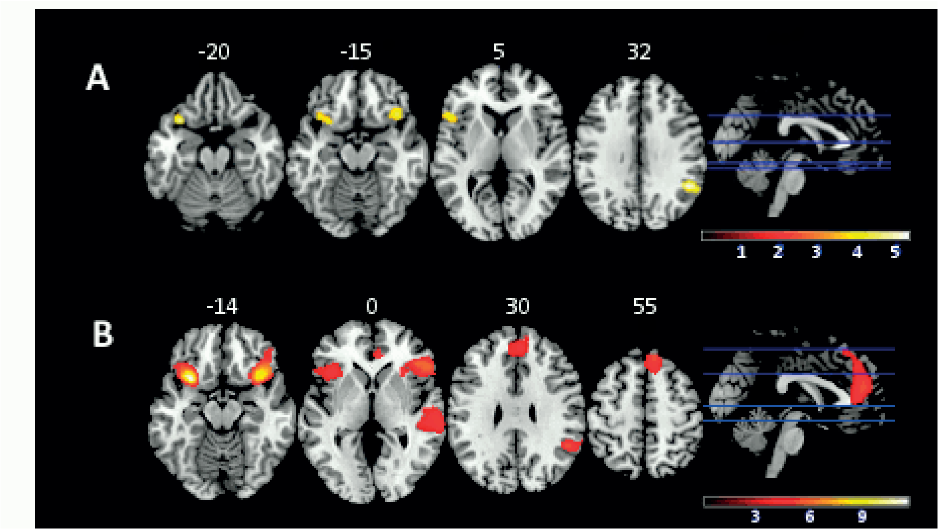


Figure S1: Significant effects in the A) Informative indirect replies > Direct replies contrast; B) Face-saving indirect replies > Direct replies contrast displayed on axial slices (z coordinate levels in millimeters).

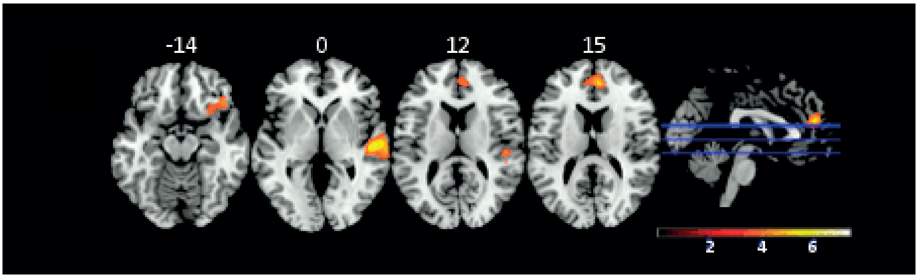


Figure S2: Significant effects in the Face-saving indirect replies > Informative indirect replies contrast. Effects are displayed on axial slices, (z coordinate levels in millimeters).

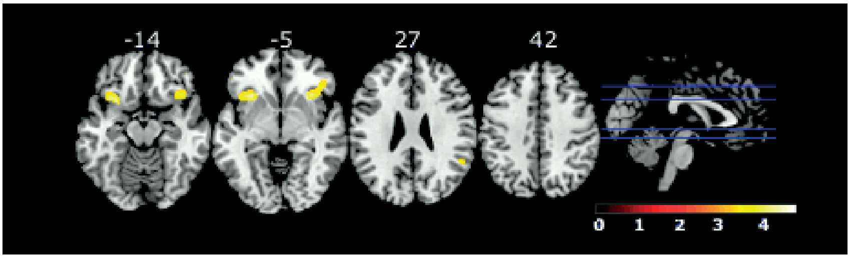


Figure S3: Significant effects for the conjunction (Conjunction Null Hypothesis in SPM5) of the two indirect conditions. Consistently with the main results, there was activation in the bilateral IFG extending to insula. In addition, there was a trend towards right temporo-parietal junction activation, it did not survive the multiple-comparison correction threshold. Effects are displayed on axial slices, (z coordinate levels in millimeters).

Anatomical region	Global and local maxima			BA	Cluster size	P-value (cluster-level FEW corrected)
	x	y	z			
Conjunction effect of the two indirect conditions						
L Inferior frontal gyrus (p.orbitalis)	-36	18	-18	45/47	442	0.004
L Anterior insula	-30	20	-6	15/49		
R Inferior frontal gyrus	42	32	-4	47/49	291	0.023
R Anterior insula	34	20	-8	15		

Table S4: Areas which were commonly activated by the two indirect conditions, as revealed in a conjunction analysis (Conjunction null-hypothesis in SPM5). Reported values are thresholded at 0.001 uncorrected, voxel level FWE corrected.

CHAPTER 3: INDIRECTNESS IN ADDRESSEES AND OVERHEARERS

A job interview in the MRI scanner: how does indirectness affect addressees and overhearers?

Jana Bašnáková[1,2], Jos van Berkum[4], Kirsten Weber[1,5], Peter Hagoort[1,3]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, NL; [2] Institute for Experimental Psychology, SAS, Slovakia; [3] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, NL; [4] UiL-OTS, Department of Languages, Literature and Communication, Utrecht University, NL [5] Harvard Medical School, Boston, USA

Abstract

In using language, people not only exchange information, but also navigate their social world – for example, they can express themselves indirectly to avoid losing face. In this functional magnetic resonance imaging study, we investigated the neural correlates of interpreting face-saving indirect replies, in a situation where participants only overheard the replies as part of a conversation between two other people, as well as in a situation where the participants were directly addressed themselves. We created a fictional job interview context where indirect replies serve as a natural communicative strategy to attenuate one's shortcomings, and asked fMRI participants to either pose scripted questions and receive answers from three putative job candidates (addressee condition) or to listen to someone else interview the same candidates (overhearer condition). In both cases, the need to evaluate the candidate ensured that participants had an active interest in comprehending the replies. Relative to direct replies, face-saving indirect replies increased activation in medial prefrontal cortex, bilateral temporo-parietal junction (TPJ), bilateral inferior frontal gyrus and bilateral middle temporal gyrus, in active overhearers and active addressees alike, with similar effect size, and comparable to findings obtained in an earlier passive listening study (Bašnáková et al., 2014). In contrast, indirectness effects in bilateral anterior insula and pregenual ACC, two regions implicated in emotional salience and empathy, were reliably stronger in addressees than in active overhearers. Our findings indicate that understanding face-saving indirect language requires additional cognitive perspective-taking and other discourse-relevant cognitive processing, to a comparable extent in active overhearers and addressees. Furthermore, they indicate that face-saving indirect language draws upon affective systems more in addressees than in overhearers, presumably because the addressee is the one being managed by a face-saving reply. In all, face-saving indirectness provides a window on the cognitive as well as affect-related neural systems involved in human communication.

Based on:

Jana Bašnáková, Jos van Berkum, Kirsten Weber, & Peter Hagoort (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers?. Neuropsychologia, 76, 79-91.

1. Introduction

Language is a powerful discrete combinatorial coding system, a verbal sign system that allows people to communicate very precisely about a potentially infinite amount of things. However, that does not mean that language comprehension is just a straightforward “code cracking” process, where word meaning is combined according to the rules of grammar. Pragmatic analyses (Clark, 1996; Grice, 1975; Levinson, 2006; Tomasello, 2008) have made it very clear that language comprehension *always* involves a range of inferential processes whereby the linguistic signs are 'contextualized', i.e., interpreted in terms of their specific context, and, above all, the likely intentions of the current speaker. Inferences about what the speaker might mean are needed at various levels of the comprehension process. We need to think about the speaker to resolve reference, in order to work out what expressions such as "I", "today", or "this paper" refer to in a particular utterance, and along the way fixate the relevant meaning of "paper". We also need it to work out things that, although not explicitly said, are conversationally implicated by the speaker in a given situation, such as when "there's a garage around the corner" is meant to also convey "and it's open now, so you can get some gas there" (Grice, 1975).

Although lagging substantially behind cognitive neuroscience research on the code-cracking aspects of language use, research on the neural substrate of this inferential side of language comprehension is now picking up speed (e.g., see Bambini & Bara, 2012; Hagoort & Levinson, 2014; Hoeks & Brouwer, 2014; Van Berkum, 2009, 2010 for reviews). One useful research strategy is to study communicative inference in situations where language itself cannot be used (e.g., Noordzij et al., 2009; Stolk et al., 2013), or where the language used is referentially ambiguous (e.g., Nieuwland & Van Berkum, 2008; Van Berkum, Brown, & Hagoort, 1999). Another fruitful approach is to study the comprehension of verbal utterances in contexts where the ultimate intended speaker meaning is very different from what is explicitly said, such as when speakers use irony (Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012), or when they are indirect (e.g., Bašnáková, Weber, Petersson, Van Berkum, & Hagoort, 2014; Jang et al., 2013; van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). Frequent activations in areas associated with 'mentalizing' or theory-of-mind (ToM) in these studies corroborate the idea that language comprehension will also typically require inferences about the speaker's perspective and intentions.

Indirectness in language is particularly interesting, because apart from providing a window on the inferential, cognitive perspective taking aspects of language comprehension (what is the speaker *really* saying?), it also provides a window on the social and therefore more affective aspects of language use. Indirectness allows us to negotiate our social identities (Holtgraves, 2002), to interact with each other more strategically, or tactfully, than we would otherwise be able to (Lee & Pinker, 2010; Pinker, Nowak, & Lee, 2008). When used judiciously, indirectness can help us get what we want without running the risk of losing face ("Care for a last drink at my place?") or of ending up in jail ("Nice store you got there, would be a shame if something happened to it"). Related, and foregrounded in research on politeness (e.g., Brown & Levinson, 1987; Goffman, 1959), indirectness can be used to make sure that *others* don't lose face ("Your draft paper does have room for improvement"). Because all of this, linguistic indirectness is an ideal domain within which to explore the interactive, inherently interpersonal and affective machinery that underlies our everyday use of language (Van Berkum, 2015; Levinson, 2006).

In a recent fMRI study (Bašnáková, Weber, Petersson, Van Berkum, & Hagoort, 2014), we examined the neural substrate of interpreting indirectness by having listeners overhear indirect utterances that involved such face-saving social navigation. An fMRI participant would, for example, hear a conversation between two friends in which one just gave a talk and asked the other whether he liked it, upon which the reply was a friendly pronounced "It is hard to give a good presentation". Compared to overhearing the same utterance in a condition where it served as a much more *direct* reply ("How is it to give a good presentation?"), such face-saving indirect replies increased activity in three important sets of brain regions. First, as predicted from pragmatic analyses of speaker meaning interpretation, working out the implicit meaning, or 'conversational implicature' (Grice, 1975), of indirect replies increased activity in core regions of the ToM network involved in cognitive perspective-taking: medial prefrontal cortex and right temporo-parietal junction (TPJ). Second, face-saving indirectness increased activation in mainly right-hemisphere temporal and frontal regions that have also been implicated in discourse-level language processing, possibly reflecting the increased inferential complexity, situation modeling and/or working memory demands associated with more complex discourse. And third, consistent with the above analysis of the social utility of indirectness, face-saving indirect replies also elicited stronger activation in insular and anterior cingulate cortex (ACC), potential signs of more *affective* perspective-taking (e.g., empathy), or of some other affective stance (e.g., sympathy, respect, disapproval) towards the speaker or his/her addressee. The aim of the current study is to replicate these fMRI findings, and to extend them to importantly different, more dynamic arenas of language use.

In everyday conversations, utterances are typically designed for particular addressees and relevant to their lives. In the vast majority of cognitive neuroscience studies on language, however, participants are presented with utterances that are completely irrelevant to their own concerns, and usually also not really addressed to them. This was also the case in our previous study, where indirect replies were embedded in dialogs by other people and about other people; dialogs which the participant was simply overhearing as an impartial listener. Even though this might not be a cause for concern in examining low-level syntactic or single-word processing, for research on the neural correlates of pragmatic interpretation, it is important to establish to what extent the findings obtained with such passive overhearing generalize to more typical communicative situations where language is more directly relevant to the listener. In real-life encounters, relevance might be achieved in several ways: either because information conveyed by an utterance informs the listener's next action (language as action rather than product; Clark, 2006), or because the listener him- or herself is directly addressed by the speaker (see Schober & Clark, 1989). Such distinctions seem to be particularly important for face-saving utterances that involve interpersonal, social navigation.

In the current fMRI study, therefore, we explore whether the neural systems that increased their activity in response to *passively overheard* face-saving indirectness (Bašnáková et al., 2014) show a similar increase in two more engaging situations: one where the face-saving indirect utterances are more actively overheard by the fMRI participant because the information is needed for a later decision (*active overhearer* condition), and one where when the indirect utterances are not only decision-relevant but also addressed to the fMRI participant him- or herself (*active addressee* condition). If our previous indirectness effects generalize to these two other arenas of language use, this would be good news for research with more passive language comprehension paradigms. However, it is not at all inconceivable that in such more engaging linguistic interactions, the use of face-saving indirect replies elicits partly different indirectness effects. We formulate more specific predictions after explaining the experimental paradigm.

1.1 The Job Interview paradigm

We created an experimental setting that provided a maximally natural context for face-saving indirect replies – a mock job interview. In a job interview, applicants can and will often use indirectness as an efficient strategy to make the best possible impression, for example attenuate the effect of potential 'shortcomings', such as not having particular experience or skill. After all, it is usually more strategic to reply to a question about whether you have some vital qualification for the job with, say, an indirect "I'm planning to take a course this summer" than with a direct "no, I don't have that skill". In the study, each fMRI participant received various job ads, together with a list of designated questions to be asked about the applicant's skills, background and experience. In the *active addressee* condition, the participant acted as an interviewer who posed the predefined questions to three "job candidates", and, critically, who therefore also was the addressee for each reply. In the *active overhearer* condition, the participant overheard *another* interviewer posing the same questions to these job candidates, and therefore also overheard each of the critical replies given to that interviewer. In both conditions, the task of the fMRI participant was to evaluate each candidate in order to choose the best one for the job. Thus, even when overhearing an interview conducted by somebody else, the participant should still be interested in going 'beyond the language given', and interpret the candidate's replies up to the deepest possible level.

As in our previous study, the main comparison was between processing of the indirect replies and that of the direct replies, the latter serving as a baseline, for example[3],

(1a) Q: "Are you fluent in any foreign languages?"

R: "I am planning to take a language course this summer" (indirect reply)

(1b) Q: "What are your plans after graduation?"

R: "I am planning to take a language course this summer" (direct reply)

---

[3] More examples of direct and indirect question-reply pairs can be found in Table 2.

Importantly, the directness or indirectness of a reply was solely determined by the nature of the question that preceded it in the interview, so that the critical replies were identical at the word and sentence level, and any differences in activation would thus reflect the processing consequences of indirectness itself. In addition, in order to maintain tight experimental control over relevant psycholinguistic variables, the candidates' replies in both active addressee and active overhearer conditions were pre-recorded and played back to the participants (see Hoeks, Schoot, Neijmeijer, & Brouwer, in preparation, for a comparable approach).

What do we predict for the differential effects of indirectness in active overhearers and active addressees? First, note that the study is not designed to test for *general* 'main effects' of overhearing vs. being addressed. Our key interest is in the neural substrate of inferential processing as well as the associated socio-affective processes in language comprehension, which we aim to make visible by comparing the response to face-saving indirect replies with that to baseline direct replies. Specifically, the aim is to see whether prior differential *indirectness effects* obtained in passive overhearers (Bašnáková et al., 2014) generalize to active overhearers, and whether findings obtained with active overhearers generalize to active addressees. It is critical to keep this differential logic in mind, and to realize that in this logic -- as in any subtraction logic -- the absence of an indirectness effect in some region of interest does not mean that this region is not engaged by both direct and indirect replies. After all, although presumably to a smaller degree than indirect ones, even our *direct* replies should require some inferencing in order to arrive at the speaker's intended meaning (Grice, 1975). And in our paradigm, also *these* utterances are part of a situation that requires the understanding and evaluation of the speaker's social moves, qualifications, and character.

For each of the two active participant roles (overhearer, addressee), we organize our predictions in terms of the three sets of processes foregrounded by face-saving indirectness in our earlier passive overhearer study (Bašnáková et al., 2014): cognitive perspective taking, other cognitive processes required to handle complex discourse, and affective processing. We also exploit a pragmatics distinction that is highly relevant for thinking about face-saving indirect replies: the distinction between the speaker's referential and social intentions (Tomasello, 2008). Referential intentions involve what a speaker wishes to draw attention to, and what listeners typically analyze in terms of "what is being talked about", i.e., a situation model (Zwaan & Radvansky, 1998). Social intentions involve what a speaker wishes to achieve, by talking about that something, at the level of social actions (Tomasello, 2008): what is it that the speaker ultimately wants a listener to do, know, or feel?

1.2 Predictions for active overhearers.

Cognitive perspective taking. Just like passive overhearers, indirectness should lead active overhearers to engage in additional cognitive perspective taking relative to direct replies, e.g., to work out that in reply to the fluency question in (1a), the utterance “I am planning to take a language course this summer” is also implicitly conveying "no, I am not fluent yet". We therefore expect core nodes of the mentalizing network to show up, at least those observed in our earlier study (mPFC and TPJ), and possibly also other ones (such as the precuneus, e.g., Schurz, Aichhorn, Martin, & Perner, 2013; Van Overwalle & Baetens, 2009; Mar 2011). Whether the *size or extent* of the indirectness effect in this network will be larger or smaller than for passive overhearers is difficult to predict. In the job interview setting, the indirect reply may well induce additional cognitive perspective taking to construe currently important aspects of the speaker's *social* intention, e.g., that this candidate prefers to downplay absent qualifications, and may even want to fool the interviewer. Furthermore, indirectness-induced mentalizing can extend to the *interviewer's* thoughts ("I wonder if she will see through this reply..."). The decision task given to active overhearers may also cause them to be more engaged across the board than in a more passive listening paradigm. If this causes all task-relevant neural systems to operate with increased 'gain' or 'intensity', comparable to what can be observed under attention-grabbing conditions in visual perception research (e.g., Vuilleumier, 2005) the differential effects of indirectness could be larger in our active overhearers. On the other hand, it is not necessarily the case that overhearers operate with reduced gain, as this also depends on whether what they are overhearing is interesting.[4] In all, although there are various good reasons to predict a sizeable indirectness effect in the

[4] Note that, although passive overhearing is usually not overly 'engaging' in the average psycholinguistic experiment, this is by no means a *necessary* implication. Consider fictional narrative, the invented stories in films, TV-series and novels that allow us to overhear (imagined or physically enacted) conversations between protagonists, and as such provide us with *massively* engaging, and often emotionally moving experiences on a virtually daily basis (Boyd, 2009; Gottschall, 2012; Mar & Oatley, 2008). In research on how people process all that fiction, central concepts such as "transportation" (e.g., Green & Brock, 2000), "identification" (e.g. Cohen, 2001), "narrative engagement" (e.g., Busselle & Bilandzic, 2008), or "absorption" (e.g., Slater & Rouner, 2002) testify to the possibility that 'passive' overhearing can be highly involving.

mentalizing network of active overhearers, the effect is not necessarily larger than that observed for passive overhearers.

Other discourse-level cognitive processes. Relative to direct replies, indirect replies should not only increase the cognitive perspective taking load in active overhearers. For the same reasons as above, they also bring additional complexity at the level of inferring and dynamically representing a more complex situation model (e.g., listeners need to model not only that the candidate will perhaps take a language course this summer, but *also* that she is not fluent yet), a more complex model of the social situation (e.g., the candidate is evading the question), and perhaps the character of the candidate (e.g., she is trying *too* hard to get the job…). We predict that these additional discourse processing requirements will additionally engage bilateral temporal and frontal regions implicated in discourse-level cognitive processing, such as inferior frontal gyrus as well as anterior, middle and superior temporal lobes (e.g., Ferstl, Neumann, Bogler, & von Cramon, 2008). As before, whether the extent of this differential activation will be larger or smaller in active as opposed to passive overhearers is difficult to predict, in part because the materials passively overheard in our earlier study (Bašnáková et al., 2014) are entirely different from the current materials.

Affective processes. In our previous study (Bašnáková et al., 2014), face-saving indirect replies increased activity in the right anterior insula and ACC in passive overhearers relative to more direct replies, possibly indexing more *affective* perspective-taking (e.g., empathy, Bernhardt & Singer, 2012; Decety & Lamm, 2006; Fan, Duncan, de Greck, & Northoff, 2011) or some other affective stance. In the current paradigm, we also predict indirectness-induced additional activity in affect-related areas, albeit not for exactly the same reasons. Previously (Bašnáková et al., 2014), participants were passively overhearing a conversation wherein other people used indirectness to navigate some distant fictitious social world. In the job interview setting, however, the participant is evaluating the candidates, and should therefore have a natural interest in the information supplied by them, in the straightforwardness and truthfulness of their replies, and in what all that says about their character. Face-saving indirect replies can therefore easily lead to increased activity in affect-related areas (see Van Berkum, 2015, for a general analysis of how affect can permeate language comprehension). Most obviously, via appraisals such as "he's trying to fool us", evasive indirect replies can lead to such affective responses as, for example, irritation (what an annoying person), competitive playfulness (you won't win this way), respect (he's playing it smart) or compassion (how sad that he feels he needs to do this) -- all the usual things that govern interpersonal interaction at the social level in everyday situations, too. Considering the diversity of specific affective responses in the current paradigm, we focus our predictions for an indirectness effect in brain areas associated with general emotional salience, notably anterior insulae and ACC (Barrett & Satpute, 2013). Whether the extent of affective responding in these areas is larger or smaller in active overhearers than in passive overhearers is impossible to predict, because of the many differences involved in this cross-study comparison.

1.3 Predictions for active addressees.

Cognitive perspective-taking and other discourse-level cognitive processing. Just like active overhearers of a job interview, active addressees of an indirect reply in that setting need to do the groundwork of additional cognitive perspective-taking, to work out the conversational implicature (e.g., "no, not fluent yet"), and to construe currently important aspects of the speaker's *social* intention (e.g., "this candidate prefers to downplay absent qualifications, and may even want to fool me"). Furthermore, all such additional cognitive perspective-taking inferences lead to additional complexity to be represented in the situation model of what is being talked about, as well as in a model of the social situation and the candidate. We thus predict that face-saving indirect replies lead to increased activation in the same mentalizing and discourse-level cognitive processing related areas as predicted for active overhearers (and observed for passive ones).

Affective processes. Just like active overhearers in the job interview setting, active addressees are evaluating the candidates, and should therefore have a natural interest in the information supplied by them, in whether replies are straightforward and truthful, and in what this says about the candidate at hand. We therefore predict that face-saving indirect replies will elicit increased activity in affect-related areas such as insula and ACC, for the same reasons as with active overhearers. However, here we can also make a strong prediction about the size of the indirectness effect, because appraisals such as "he's trying to fool me with this reply" or "he seems to think that people like me can easily be misled" get very personal: it is the fMRI participant him- or herself that is being fooled, misled, or otherwise manipulated by the job applicant. We therefore predict that, at least for this reason, evasive indirect replies will be more emotionally salient for active addressees than for active overhearers, and will as such lead to stronger increases of activity in the emotional salience network (Barrett & Sadpute, 2013), as well as perhaps in other affect-related brain areas.

Taken together, the general prediction is that the pattern of findings obtained with passive overhearers (Bašnáková et al., 2013) will port to active overhearers and active addressees in a job interview setting, for the simple reason that the same types of additional processing needs to be done. So, relative to straightforward direct replies, face-saving indirect replies should lead to increased activity in mentalizing areas (at least mPFC and rTPJ, and possibly precuneus), in fronto-temporal areas dealing with other discourse-level cognitive complexity, and in affect-related areas (at least insula and ACC). Although arguments concerning additional involvement and 'increased gain' are tempting to make, we refrain from making strong specific predictions about the size and extent of those indirectness effects, except for one particular case: because the active addressee in the interview setting is the very person being 'manipulated' by the socially navigating speaker, we expect face-saving indirect replies to elicit stronger affect-related activations in addressees than in active overhearers.

2. Materials and Methods

2.1 General design

To create a situation where posing questions and getting direct and indirect replies would be natural, we set up a mock *job interview* experiment. Each fMRI participant interviewed three job candidates for different jobs (teacher, manager, office assistant, researcher) and on the basis of their replies made a decision about whom to hire (a short example interview is in *Appendix 1*). The participant was told that the three job candidates they were to interview were role-playing participants who already completed a similar session with another fMRI participant 'interviewer' the week before, and that he or she would now conduct a live second interview with them (active addressee condition), as well as listen to the earlier interviews (active overhearing condition). The fMRI participant was also told that the live interviews with each of the three job candidates had to be conducted over the intercom to those candidates in an adjacent room, "to make sure that any accidental differences in attractiveness between candidates would not influence the participant’s choices". Each job interview was structured by means of a predefined set of questions, designed to help in choosing the best candidate.

This allowed us to maintain tight experimental control over linguistic aspects of the critical replies across the two conditions: both replies that were given in the 'live' interview (active addressee condition) and in the “recorded” interview (active overhearer condition) were actually recorded earlier by the same native Dutch speakers, now played back at the right time by the experimenter. After the interviews were done, we excluded those fMRI participants who did not believe that they were talking to live people in the active addressee condition, from the analysis.

2.2 Materials

We constructed 120 replies and 240 questions: each reply served either as a direct reply or an indirect reply when preceded by its corresponding “direct” or “indirect” questions. The indirect questions were yes/no question, so that the listener was forced to interpret the reply as either a ‘yes’ or a ‘no’, which required inference from the actual reply. The direct questions were open-ended (why…? what…? etc.). Example stimuli are in Table 2, and the full set of written Dutch materials can be obtained by request from the first author.

The replies were recorded by three female Dutch native speakers (a third each) and all the questions were recorded by another female Dutch native speaker who acted as the interviewer in the overhearer condition. In order to preserve the dynamics of an interview, the recordings were carried out as if the two speakers were actually in an interview, i.e. the reply always followed a question. The speakers were instructed not to read the QA pairs during recordings, but to say them out of their memory. The final set of stimuli was edited in Praat (Boersma & Weenink, 2009), adjusting the volume and cutting the question-answer pairs into individual conversational turns. Questions started from the beginning of the first word and ended just after the last word, and replies were cut from the end of the preceding question, with the pre-utterance pause included.

Even though the direct and indirect replies were identical at the word- and sentence- level across conditions, we took great care in attenuating any possible differences between the conditions other than the replies’ differences in implied message. We equated the direct and indirect questions on:

(i) length in the number of words (mean direct question = 12.81, mean indirect question = 12.86); (ii) the extent of direct semantic priming – i.e. how many content words from the question were repeated in the reply; (iii) loudness of the critical utterance across conditions; (iv) semantic similarity of the question and the reply. Semantic similarity was compared by means of latent semantic analysis (Landauer, Foltz, & Laham, 1998; Landauer & Dumais, 1997) (available online at http://lsa.colorado.edu). Since there are currently no reasonably large semantic corpora available in Dutch, we translated the stimuli into English and carried out the analysis in English. A term-to-term comparison of content words using tasaALL space, which corresponds to a reading level of a third-grade college student, yielded similar mean semantic similarity values (SSVs): direct = 0.65, indirect = 0.66.

We also carried out two pre-tests on different aspects of the stimuli. Firstly, 5 Dutch college-aged native speakers judged the QA pairs for "acceptability", which was defined as "the extent to which you find the reply acceptable" within the context of a job interview. Participants were supposed to judge the acceptability on a 5-point scale. We changed the QA pairs which yielded low ratings and an independent Dutch native speaker judged them again, until they were all acceptable. After that, we conducted an *Indirectness pretest*, where another group of 22 native Dutch college students judged, on a 7-point scale (1 = completely direct, 7 = completely indirect), to what extent the critical utterance was a direct or an indirect reply to the preceding question. There was a significant difference between the average rating for direct (2.84 (SD 0.91)) and indirect (5.65 (SD 0.66)) replies (paired-samples t-test, $p<0.001$). In order to make the difference even more pronounced, we further changed items whose average difference in ratings for the direct and indirect version was less than 2.5 points.

Although the critical direct and indirect utterances were identical in their wording, there were small differences in the duration of each reply due to prosody, and in intonation. Mean duration of direct replies was 3.35 s (SD 0.80), and of indirect replies 3.62 s (SD 1.03), a statistically significant difference ($p< .001$). Also the pre-utterance pause was, on average, longer for the indirect replies. We deliberately chose not to make these acoustic delivery properties uniform or to use a single token of the utterance in both conditions, as this would compromise the naturalness of our materials.

DIR: What is, in your opinion, the biggest shortcoming of internships? IND: During your internship, have you worked independently on any big projects?	*Interns do not usually get really important projects to work on.*
DIR: What inspired you to go into people management? IND: Have you followed any certified courses on leadership?	*I've read a couple of very inspiring books about leadership.*
DIR: How and when did you acquire basic command of German? IND: You say that you are fluent in German. Have you followed any certified courses?	*I've read a lot of books in German during college.*
DIR: What was the biggest hurdle in getting funding for your PhD? IND: Have you received any grants or scholarships during your studies?	*The competition for scholarships in my field is extremely harsh.*

Table 2. Examples of direct (DIR) and indirect (IND) question-answer pairs used in the Job interview study. The critical event was always only the reply. The stimuli were presented in Dutch.

We also constructed 147 filler items with mostly open-ended questions (such as wh-questions) and direct replies. Filler items were purposefully imperfect, with dysfluencies, repairs, false starts, etc., to convey the impression that the candidates were speaking in real time.

We created two stimulus lists with the final stimulus set – half of the participants were presented with list A in the overhearer condition and list B in the addressee condition, and *vice versa* (so that a participant would never hear the same reply twice). Each list contained 30 direct and 30 indirect replies and the rest were fillers. Once again, even though the target utterances – the replies – were identical for both direct and indirect conditions, and served as their own controls, we made sure that important psycholinguistic variables were as balanced as possible in both versions of the stimulus lists for both direct and indirect conditions: lexical frequency of content words in the target utterances, number of words per direct/indirect questions and direct lexical priming (repetition of the same content words from question to reply).

Finally, we created another 60 single sentences for a "non-communicative control" condition, matched on word count (version A = 10.07 (SD 2.48), version B = 10.07 (SD 2.36)) and lexical frequency of content words (A = 3.40 (0.31), B = 3.41(0.37)). There was an "addressee" control which was presented after the Addressee part of the job interview, and an over-hearer control. In each of these conditions, participants were instructed to simply listen for meaning to a set of 30 isolated sentences, spoken by the candidates but unrelated to the interview. In the addressee condition, they were told to "listen to the candidates reading these sentences for you", and in the overhearer condition, they were told to "listen to these sentences recorded at a previous session with the candidates". This provided us with a way to probe the addressee/overhearer distinction at a very basic level of somebody speaking directly to you right there and then, as opposed to hearing a recording, without a richer social context that turns it into useful communication.

2.3 Procedure

We communicated the study as a social decision-making study to the participants. We took several measures to make the "addressee" manipulation convincing. Starting with the recruitment information, we pretended to look for 4 persons per session, with one playing the role of the interviewer (the actual participants) and the rest the role of "job candidates". Before the session, we emailed the fMRI participant that they were chosen for the role of the interviewer, asking them not to come earlier to their appointment, as to not meet the rest of the participants ("job candidates") at the door. In addition, before the addressee part of the session, the experimenter pretended to call her colleague in the adjacent room to find out whether the three job candidates were ready, and she then proceeded to a "sound check", in which each of the candidates and the actual participant tested their microphones. The candidates' voices for the sound check were recorded and manually triggered by the experimenter. As for the content of the candidates' replies, we told the fMRI participant that each candidate had a rough outline of what his character was like (skills, knowledge, experience) but was free to improvise to some extent. However, candidates could not lie, e. g. say that they have certain skills when in fact they don't.

Both addressee and overhearer parts of the interview took place on one day, with their order of presentation counterbalanced across participants. Each part was followed by a short control run, where the fMRI participant heard 30 sentences from the candidates (10 each). After the participants came out of the MRI scanner, they took part in an exit interview.

Each interview (in both addressee and overhearer conditions) followed the same pattern: first, a short job advertisement was projected on the screen, stating the requirements for a specific job. There were four positions in each part (teacher, manager, researcher and office manager). When the participant was ready, the interview started. There were 11 or 12 questions for each candidate, which the fMRI participants either read out loud themselves from the computer screen (addressee condition) or listened to them through headphones and saw them projected on the screen at the same time (overhearer condition). The participants were asked to press a button with their right index finger when they were finished with reading or listening, to indicate that they were ready for the answer. The button-press triggered the recorded reply after a 1.5 s delay. After each candidate answered their set of questions, the fMRI participant was to indicate on a 7-point scale how much he or she liked the candidate, how much he or she thought this candidate was suitable for the job, and how much this candidate was similar to him or her. Then, the next candidate followed. When all three candidates were interviewed (either by the participant or by another interviewer), the fMRI participant was asked to choose the best one for the job and then verbally indicate the reason for this choice. He was instructed to say whatever he finds relevant to support his choice and in the addressee condition, he was assured that the candidates cannot hear his judgment. This was done in order to make the two conditions as similar as possible. After the decision, the next job advertisement and round of interviews followed.

The composition of each experimental trial was as follows: a fixation cross was projected in the middle of the screen for a variable time interval between 1.3 and 3.3 seconds. Then, a question was displayed, terminated by a button-press. The length of the question duration was variable, as it depended on how fast participants read it (as addressees) or how long it took them to comprehend the spoken questions (as overhearers). A fixation cross appeared on the screen for a variable time interval between 1.6 and 4.6s, which included a pre-utterance pause before the replies. The fixation cross stayed on the screen during the critical event – the reply – which was presented auditorily, followed by 1.5 seconds of silence.   The replies lasted on average 3.49s (SD 0.93, range 1.76 – 7.99). The entire session lasted for 2 – 2.5h.

After the session, we conducted an exit interview with the fMRI participants. Here, we asked them whether they felt any subjective differences between interviewing the candidates themselves and listening to another person interviewing them (e.g. whether it was more difficult to make their decisions), whether they noticed any "evasive" replies, understood all the replies, and found any of the candidates more likeable than others. We also asked them to tell us whether they noticed anything strange and gave them the opportunity to ask us anything they wanted to know about the entire study. As the last "credibility check" of our addressee manipulation, we asked them whether they now wanted to meet the candidates in person. Only a few participants wanted to – to those, we pretended to look for the candidates and then told the participants that they've already left. Out of the 30 fMRI participants, 8 indicated in the exit interview that they did not believe they interacted with real people. These were debriefed about the purpose of the study immediately. The rest was debriefed via e-mail after the study was finished.

2.4 Participants

30 students participated in the experiment for money or course credit and 20 of them were included in the final analysis (4 male, mean age 20.8, SD 2.6). Eight participants were excluded because they did not believe in the "addressee" manipulation and two because of other reasons (one had difficulties reading and comprehending the questions, one did not finish all the runs in the required order).

2.5 fMRI Data Acquisition

Participants were scanned in a Siemens 3-T Tim-Trio MRI-scanner using an 8 channel surface coil. Functional images were acquired using an EPI multi-echo sequence (Poser, Versluis, Hoogduin, & Norris, 2006) in an ascending order. The repetition time (TR) was 2.35 s and each volume consisted of 36 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 4 different echo times (TE1) = 9.4ms, (TE2) = 21.2ms, (TE3) = 33ms, (TE4)= 45ms. Flip angle was 90°. A whole-brain high-resolution structural T1--weighted MPRAGE sequence was performed to characterize participants' anatomy (TR = 2300 ms, TE = 3.03 ms, 192 slices with voxel size of 1 mm3, field of view = 256).

2.6 fMRI Data Analysis

The functional magnetic resonance imaging (fMRI) data were preprocessed and analyzed using Statistical Parametric Mapping (SPM8, fil.ion.ucl.ac.uk/spm/). The first pre-task 30 volumes were acquired and used for weight calculation of each of the echoes. The functional echo-planar imaging-BOLD images were then realigned and slice-time corrected. The resulting functional images were co-registered 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 4 types of the target utterances (direct reply addressee, indirect reply addressee, direct reply overhearer, indirect reply overhearer), which were defined from the start of speaking (not including the pre-utterance pause), including the entire reply and 1.5s of silence at the end of it. Each condition included 30 trials. We also modeled the onset and duration of the visually presented fixation crosses before each question as a baseline. Since the Overhearer and Addressee conditions were measured in separate runs with a break in between, the baseline included fixation periods from both sessions. We did this in order to account for any potential differences in the baseline between these two runs. In addition, we modelled the onsets and durations of the questions. In a separate analysis of the control sentences, we modeled onsets and duration of each control sentence and a fixation cross preceding the sentence. 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, the following images were defined for each participant and used in the second-level random effects analysis: indirect reply > fixation cross baseline and direct reply>baseline (in both overhearer and addressee conditions), as well as control sentence > fixation cross in both addressee and overhearer control conditions.

In the second-level random effects analysis, we used the contrast images of interest (reply > baseline) in a repeated-measures flexible-factorial model with Listener Status (addressee, overhearer), Indirectness (direct, indirect) and Subjects as factors. The cluster size was used as the test statistic, and only clusters significant at $p < 0.05$ corrected for multiple non-independent comparisons are reported. The initial threshold was 0.001 at the voxel level.

2.7 Region-of-interest analyses

To test our predictions about cognitive perspective-taking, additional discourse-induced cognitive processing, and affective processing with more statistical power, we conducted several Region-Of–Interest (ROI) analyses. The first one included 4 ROIs that are considered core nodes of the mentalizing or ToM network (e.g. Schurz et al., 2013; Mar, 2011): left and right temporo-parietal junction (TPJ), medial prefrontal cortex (MPFC) and precuneus (PC). The coordinates were taken from a recent meta-analysis on ToM by van Overwalle and Baetens (2009): right TPJ: x=50, y=-55, z=25; left TPJ: x=-50, y=-55, z=25; MPFC: x=0, y=50, z=20; and PC: x= 0, y= -60, x= 40. Second, to asses additional discourse-induced cognitive processing, we chose two sets of areas routinely activated in studies that go beyond sentence-level meaning comprehension: bilateral anterior temporal lobes (ATL) and bilateral inferior frontal cortex. Bilateral ATL is the most stably activated part of the extended language network (Ferstl et al., 2008) observed in text comprehension research, and was also sensitive to implied speaker meaning in another study on conversational implicature comprehension with indirect replies (Jang et al.,2013). ATL coordinates were taken from the latter study: left ATL: x =-50, y=8, z=-24; right ATL: x =54, y=0, z=-20. Coordinates for bilateral IFG, both BA 45 and 47, were taken from Hagoort et al. (Hagoort, Hald, Bastiaansen, & Petersson, 2004): left BA45 x=-44, y= 30, z= 10; right BA45: x= 44, y= 30, z= 10; left BA 47: x=-48, y=30, z=-13; right BA47: x=48, y=30, z=-13. These areas were sensitive to world-knowledge anomalies in the Hagoort et al. study, and (apart from left BA45) also involved in unifying generic world-based scenario knowledge with specific discourse-context constraints in a follow-up study (Menenti et al., 2009).

The third ROI analysis investigated the ventral salience network (Barrett & Satpute, 2013) with bilateral anterior insulae (AI) and pregenual anterior cingulate cortex (pACC). Coordinates for this analysis were taken from a resting state connectivity study by Taylor et al. (2009): L anterior insula: x= -34, y =14, z =2; R anterior insula: x = 36, y=16, z=2 ; left pACC: x = -8, y= 38, z= 19; right pACC: x= 8, y = 38, z = 19[5].

Using the MarsBaR toolbox for SPM (Brett, Anton, Valabregue, & Poline, 2002), we extracted mean signal from an 8-mm spherical ROI around each of these peak coordinates and subjected them to repeated-measures ANOVAs in SPSS (version 23) with ROI, Listener Status and Indirectness as within-subject factors.

3. Results

We report main indirectness effects and interactions with the active overhearer/addressee factor in a whole brain analysis first, and then report results of ROI analyses in the networks of interest.

3.1 Main effect of indirectness

Comparing the indirect with direct replies across both addressee and overhearer conditions (*indirect > direct*, see Fig 1 and Table 1) engaged a bilateral fronto-temporal network of areas which largely overlapped with the comparison between face-saving and direct replies in our previous study on processing indirect replies (see Fig 2c, reprinted from Bašňáková et al., 2014). In the current study, indirectness elicited bilateral activations in the inferior frontal gyri and anterior insulae; in the middle and superior temporal gyri, extending into the temporal poles. Moreover, there were bilateral activations in the angular gyri (TPJ), and superior medial frontal gyri, extending into ACC. Unlike in the previous study, the indirectness effect also activated subcortical regions in the right hemisphere, namely thalamus and pallidum.

---

[5] The emotional salience (Taylor et al., 2009) and mentalizing (van Overwalle & Baetens, 2009) sets of coordinates were first converted from the Talairach to the MNI space using the tal2mni transform script implemented in Matlab. Discourse-level processing coordinates from Hagoort et al. (2004) are reported in MNI space.

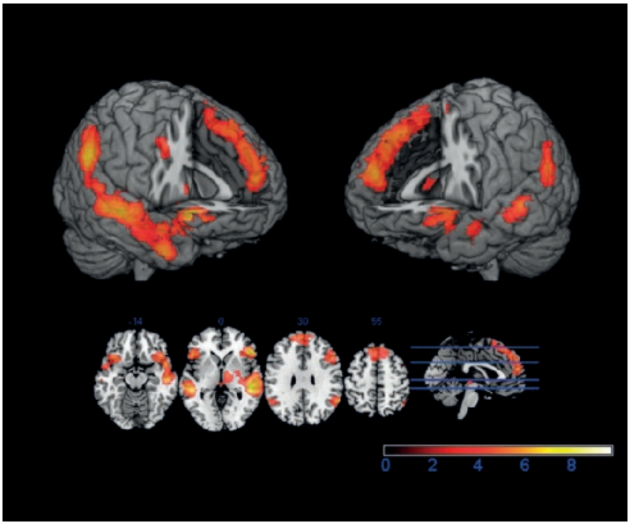


Figure 1: Brain areas activated for the Indirectness effect (contrasting indirect > direct replies) in the current study across both Addressee and Overhearer conditions. Significant effects are displayed on cortical renderings and on axial slices (z coordinate levels are in millimetres: -14, 0, 30, 55). Color bar denotes t-values (in blue).

Anatomical region	Coordinates of local maxima			Cluster size	p-value (cluster-level FWE corrected)
	x	y	z		
R middle temporal gyrus	50	-26	-6		
R inferior frontal gyrus (p. triangularis)	52	26	4		
	40	20	30		
R medial temporal pole	48	6	-34		
R angular gyrus (TPJ)	60	-56	34		
R superior temporal gyrus	60	-16	0		
R superior temporal pole	52	8	-14		
R inferior frontal gyrus (p. opercularis)	48	18	32	9242	<0.001
R precentral gyrus	44	8	48		
R thalamus	10	-10	2		
R insula	34	18	-12		
R pallidum	30	-14	-6		
R middle frontal gyrus	42	4	42		
R thalamus	12	-14	-4		
basal ganglia thalamus, extending into pallidum	0	-26	-2		
	-2	48	38		
	-4	50	34		
L medial superior frontal gyrus	-8	34	56		
	-10	52	34		
	2	56	20	3268	<0.001
	2	34	52		
	4	44	44		
R medial superior frontal gyrus	12	60	24		
	12	62	20		
	2	46	38		

L supplementary motor area	-12	10	68		
R supplementary motor area	16	28	56		
L inferior frontal gyrus (p. triangularis)	-54	20	10		
	-46	26	-2		
	-36	20	-14		
L inferior frontal gyrus (p. orbitalis)	-46	32	-4	1829	<0.001
	-40	24	-10		
L temporal pole	-38	16	-18		
	-48	2	-24		
L middle temporal	-52	4	-14		
L middle temporal gyrus	-52	-28	-6		
	-48	-36	0		
L superior temporal gyrus	-44	-26	10	1382	<0.001
	-44	-28	14		
L angular gyrus (TPJ)	-58	-58	26		
	-56	-58	40	519	0.008
L middle temporal gyrus	-54	-56	22		

Table 1. Brain areas activated in the contrast Indirect > Direct, cluster size p-value 0.001 (cluster-level FWE corrected)

3.2 Interaction between listener status and indirectness

At the level of the whole brain analysis, the processing consequences of indirectness did not significantly depend on whether the participant was an addressee (Fig 2A) or an overhearer (Fig 2B). There were no significant clusters activated for the interaction at the $p < 0.001$ threshold at both sides of the interaction (t-contrasts).

3.3 Region of interest analysis

As illustrated in Figure 3A, we did not find any evidence that critical regions implicated in cognitive perspective-taking respond differently to indirectness depending on whether the participant is directly addressed or is actively overhearing them. For this mentalizing ROI analysis, there were significant main effects of Indirectness ($F(1,19) = 14.87$, $p = 0.001$) and Listener Status ($F(1,19) = 7.79$, $p = 0.012$) but no interactions ($F(1,19)=0.23$, $p=0.635$; Listener Status with Indirectness and ROI: $F(3,57)=6.84$, $p = 0.089$). Thus, although there was greater activity for indirect than direct replies in all four regions of interest, the size of this indirectness effect in these areas did not reliably depend on whether the participant was overhearing or being addressed.

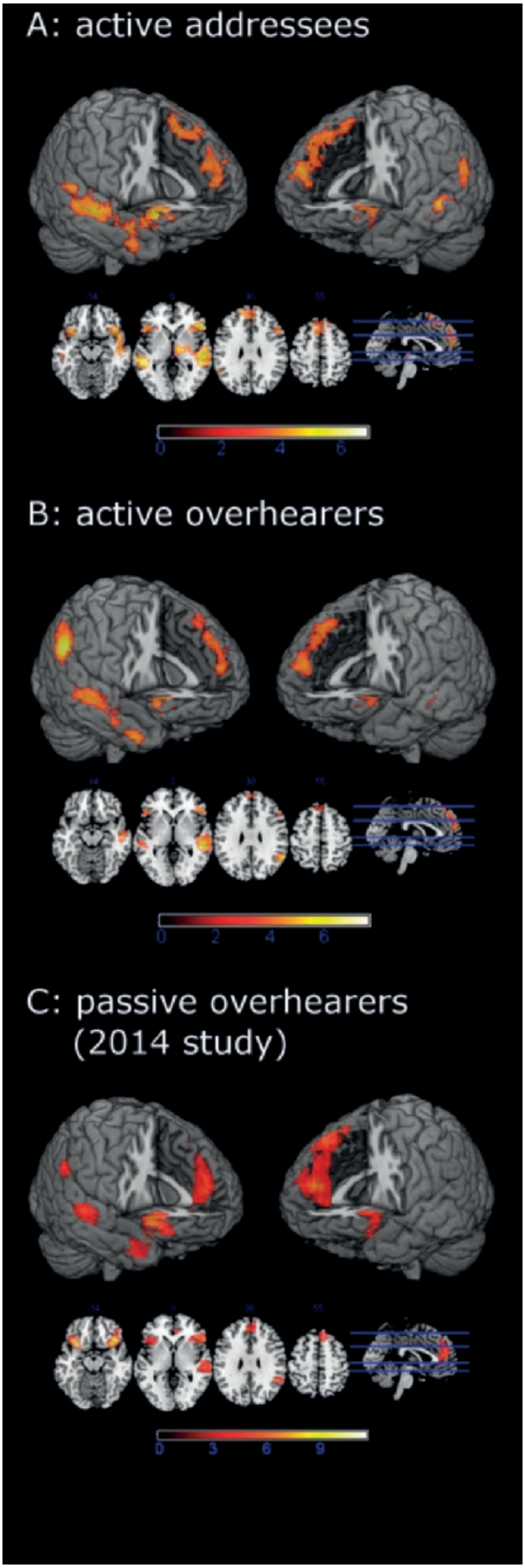


Figure 2: Brain areas activated for the Indirectness effect (indirect face-saving > direct replies) in (A) active addressees and (B) active overhearers, of the current study, and (C) passive overhearers in the Bašnáková et al. 2014 study. Even though graphical comparison seems to suggest a lateralization difference in the temporo-parietal area, there is no interaction at the whole brain level. Significant effects are displayed on cortical renderings and on axial slices (z coordinate levels are in millimeters: -14, 0, 30, 55). Clusters are shown at a threshold of 0.001 voxel level and extent of >200 voxels. Color bars denote t-values (in blue).

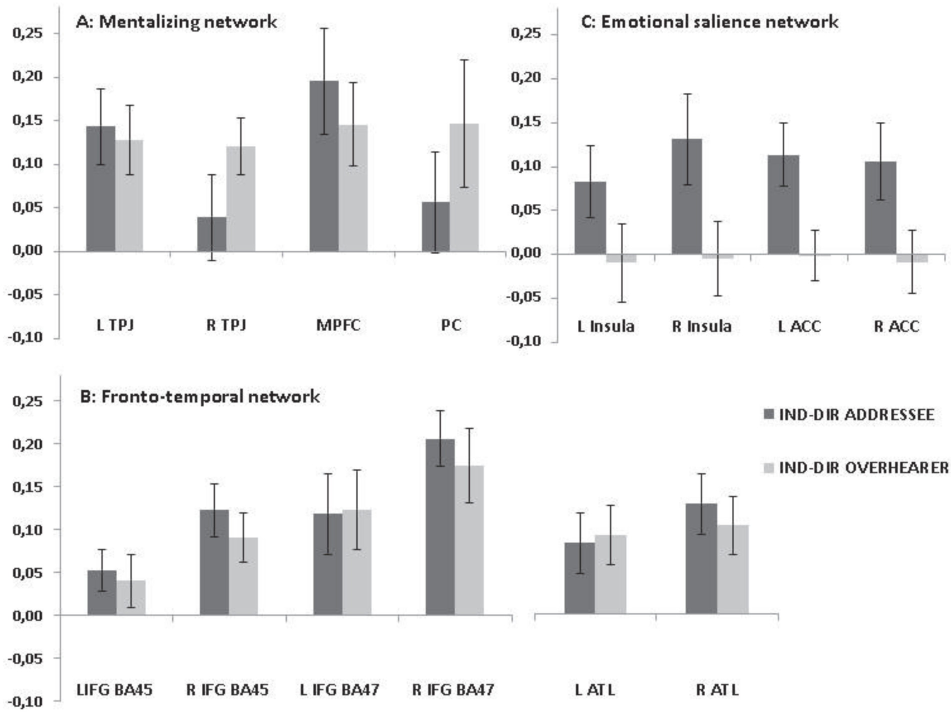


Figure 3: The effects of face-saving indirectness. Mean differential HRF signal for indirectness (Indirect > Direct), as a function of listener status (active addressee vs. active overhearer) in (A) mentalizing regions of interest left and right TPJ, medial prefrontal cortex, precuneus; (B) discourse-level cognitive processing ROIs: left and right IFG (BA45 and 47), bilateral anterior temporal lobes; and (C) the ventral affective salience network ROIs: left and right insula and left and right ACC. Displayed are the differences in average beta values between indirect replies and direct replies at each ROI. Error bars indicate SEM.

A similar pattern of results was obtained in the two ROI analyses aimed at discourse-level cognitive processing regions in the fronto-temporal network, see Figure 3B. The bilateral ATLs showed greater activation for indirect than for direct replies ($F(1,19) = 35.10$, $p < 0.001$), but the size of this indirectness effect did not reliably depend on listener status ($F(1,19) = 0.03$, $p = 0.875$), in none of the ROIs ($F(1,19) = 0.54$, $p = 0.473$). Likewise for the bilateral IFG (BA45 and BA47): greater activation for indirect than for direct replies ($F(1,19) = 34.16$, $p = <0.001$), in a way that depended on ROI (indirectness * ROI interaction, $F(3,57) = 9.62$, $p = 0.001$, after Greenhouse-Geisser correction), but again no reliable impact of listener status on the size of this indirectness effect ($F(1,19) = 0,19$, $p = 0.669$), for any of the ROIs ($F(3,57) = 0.23$, $p = 0.801$, Greenhouse-Geisser corrected).

The ROI analysis for the affective ventral salience network revealed a very different pattern of findings. Although there was a trend towards a main effect of indirectness ($F(1,19) = 4.23$, $p=0.054$) in these areas, the pattern is dominated by an interaction ($F(1,19) =6.08$, $p = 0.023$), see Figure 3C: indirectness increased activation in these areas when people were themselves addressed, but not when they were overhearing the indirect replies.

In line with the aims of our study, we focus on the *differential impact* of face-saving indirectness and its modulation by listener status. Although our design is less optimal for examining *main effects* of listener status (Addressee>Overhearer and vice versa), a report on the latter, as well as on results for a related non-communicative control condition, can be found in the Supplementary materials.

4. Discussion

We examined an ingredient of language comprehension that, although critical, has not received all that much attention from researchers in the neurobiology of language field – how listeners arrive at the real speaker meaning, beyond the literal code. Building on our prior fMRI study on pragmatic inferences behind indirect replies (Bašnáková et al., 2014), we extended the work in two directions. Firstly, we moved from a paradigm in which listeners were passively overhearing bits of conversation towards a paradigm in which listeners needed to really comprehend those bits of conversation for further decision-making, a dominant mode of language use in the real world. To that end, we created a mock job interview setting, where, the fMRI participant needed the information gleaned from direct and face-saving indirect replies to make decisions about which job candidates to hire. In addition to this active overhearer condition, our second extension was to compare the impact of face-saving indirect language use to that of the same utterances in active addressees. In the relevant condition, fMRI participants believed they were themselves interacting with the job candidate, and that direct and indirect replies were thus addressed to *them*.

4.1 Indirectness when actively overhearing

First, what is the differential impact of a face-saving indirect reply when actively overheard? We had predicted increased activation in core nodes of the mentalizing or ToM network, at least those observed in our earlier study (mPFC and TPJ), and possibly also other ones (e.g., precuneus). This is exactly what we observed. In the whole brain analysis, indirect replies induced additional activation in mPFC and right TPJ. In a more powerful region-of-interest analysis that examined the four most typically activated mentalizing regions (mPFC, bilateral TPJ and precuneus; Schurtz et al., 2014; Van Overwalle & Baetens, 2009; Mar, 2011), *all* of these predesignated areas were recruited more by indirect replies than by direct replies. This is in line with theoretical analyses in pragmatics (e.g. Grice, 1975) according to which the comprehension of what speakers *really* mean requires a consideration of their communicative intentions, which inevitably involves taking their perspective into account.

The mentalizing network highlighted here also shows up in other studies on the comprehension of speaker meaning in overhearers, albeit not always to the same extent (e.g., Bašnáková et al., 2014: mPFC and right TPJ; Van Ackeren et al., 2012: mPFC and bilateral TPJ; see also Hervé, Razafimandimby, Jobard, & Tzourio-Mazoyer, 2013; Li, Jiang, Yu, & Zhou, 2014). All four nodes, including the precuneus, were activated in a recent study on the comprehension of speaker meaning in ironic utterances (e.g., "Tonight we gave a superb performance" said by one opera singer to another after a disastrous performance; Spotorno et al., 2012). The same four nodes were more strongly activated by referentially ambiguous pronouns (as in "When Beyoncé met Madonna she had just had a little accident at the hairdressers") than in unambiguous controls (compared to, e.g. "When Beyoncé met Prince she had just had a little accident at the hairdressers"; Nieuwland, Petersson, & Van Berkum, 2007). The latter suggests, in line with pragmatic analyses, that cognitive perspective-taking is not just needed for deriving rich conversational implicatures in indirect or ironic statements, but is also involved in more 'mundane' aspects of utterance comprehension, such as fixing the referent of a pronoun. This supports an important point we already made: although we use indirectness as a way to *uncover* the neural substrate involved in the inferential side of language comprehension, cognitive perspective-taking presumably occurs with *any* contextualized utterance, for the simple reason that listeners will want to know what the speaker is *really* talking about, and what his or her social intentions are.

People have noted that the areas implicated in mentalizing strongly overlap with nodes of the default-mode network (DMN) that is active when participants are not performing a specific task, e.g. during rest periods (e.g., Buckner, Andrews-Hanna & Schacter, 2008). However, this overlap may not be a coincidence -- it has been suggested multiple times (e.g. Mars et al., 2012; Schilbach, Eickhoff, Rotarska-Jagiela, Fink, & Vogeley, 2008) that when participants do “nothing” during rest periods, they are presumably doing precisely what is characteristic of social cognition – engaging in self-referential thinking, musing on what others’ have said or done to them, and planning future social action. Also, a more generic explanation of our findings in terms of particular task conditions being more *difficult* than others and as such pulling more resources away from the brain's "default state" (a possibility that is sometimes considered when comparing task condition effects *regardless* of whether they involve social cognition) would need to suppose that the direct reply is more difficult than the indirect reply. In all, we believe that an explanation in terms of additional cognitive perspective-taking is currently the most plausible one for our indirectness effect in TPJ, MFC and PC.

Our second prediction was that actively overheard indirectness would also increase activity in fronto-temporal areas known to be sensitive to other cognitive implications of additional complexity in rich discourse situations, such as bilateral frontal and temporal regions implicated in semantic processing, causal inferencing, working memory and executive processes (e.g., Bohrn, Altmann, & Jacobs, 2012; Ferstl, Neumann, Bogler, & von Cramon, 2008; E. C. Ferstl & von Cramon, 2002; Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006; Rapp, Mutschler, & Erb, 2012; Menenti et al. 2009). Recall that an indirect reply such as “I am planning to take a language course this summer” brings additional complexity to the 'situation model' (e.g., the listener needs to add that the candidate will perhaps take a language course this summer, *and* that she is not fluent yet), as well as to representing the candidate at a more social level (e.g., is evading the question, is trying too hard). Our findings confirm this prediction as well. In the whole brain analysis for active overhearers, indirect replies increased activation in bilateral IFG and bilateral temporal regions, results that were echoed in our more sensitive region-of-interest analysis on BA45, BA47, and ATL We return to these areas when we discuss the findings for active addressees.

Based on our earlier results as well as a conceptual analysis of the current active overhearing situation, we had also predicted an indirectness effect in nodes of the affective ventral salience network (Barrett & Satpute, 2013), notably anterior insular cortex and pregenual ACC. In the regions of interest analysis, this prediction did not hold up. This is surprising, because in a job interview setting, indirect replies can easily be assumed to lead to emotion-inducing appraisals of the speaker's evasive move (irritation, sympathy, etc.). One might be tempted to pursue the idea that active overhearers approach matters in a relatively balanced way in the current paradigm, such that they find straight answers as affectively (un)engaging as face-saving ones, without being sensitive to the potentially affective overtones of the latter. Note, however, that in the whole brain analysis, face-saving indirectness did increase activity in *other* areas commonly associated with affect, namely slightly more dorsal region of right insula.

4.2 Indirectness when being addressed

An important benefit of our current paradigm is that it also allows us to compare, within a single study and with the same participants, what face-saving indirectness does with listeners who believed they were actually *being addressed*, as opposed to active overhearing the same indirect utterances. Just like active overhearers, active addressees of an indirect reply need to do the groundwork of additional cognitive perspective-taking, to work out the conversational implicature (e.g., that “I am planning to take a language course this summer” also means "no, not fluent yet"), as well as the speaker's *social* intention (e.g., "this candidate prefers to downplay absent qualifications, and may even want to fool me"). We therefore predicted that face-saving indirect replies should in any case increase activation in core nodes of the mentalizing network of active addressees as well. This first prediction was confirmed, by increased activation for indirect replies in all ROI-targeted regions: medial PFC, bilateral TPJ, and precuneus (see Figure 3A).

Interestingly, both the whole brain analysis and the more sensitive region-of-interest analysis also revealed that in the current job interview paradigm, the core nodes of the mentalizing network do not reliably care whether the indirect utterance is overheard or directly addressed to you. All of these areas were more highly activated for indirect than for direct replies, in both the addressee and overhearer conditions, and the size of the indirectness effects in those areas did not reliably depend on whether one was overhearing or being addressed. This is in line with our idea that in comprehension, the same mentalizing groundwork needs to be done in either case. Moreover, it suggests that these components of the comprehension system do not necessarily operate with increased gain just because somebody is being addressed.

Active addressees and active overhearers also do not differ in how they respond to face-saving indirectness in fronto-temporal regions associated with other language-induced cognitive operations (e.g., bilateral IFG, bilateral MTG, right STG, and right temporal pole), neither in the whole-brain analysis, not in more targeted ROI analyses. In these latter analyses, indirect replies induced additional activations in left and right IFG (BA45 and BA47), and in left and right ATL. Furthermore, as illustrated clearly in Fig 3B, these differential effects of indirectness in active addressees were not reliably different from those in active overhearers.

The involvement of bilateral inferior frontal gyrus is consistent with the idea that the comprehension of indirect replies requires intensified consideration of the context, such as the preceding question, the wider discourse, and the social rules that we obey when talking to each other. This area has been involved in studies on both sentence (Hagoort, Baggio, & Willems, 2009) and discourse-level processing, such as making causal inferences (Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006) or in supporting semantic selection of inferential information (Mason & Just, 2011). The right IFG, specifically, has been related to the difficulty of integrating incoming information into context (e.g. Tylén, Wallentin, & Roepstorff, 2009; Wang, Lee, Sigman, & Dapretto, 2006; Tesink et al., 2009; St George, Kutas, Martinez, & Sereno, 1999).

Anterior temporal lobes are also found in many studies on discourse-level processing (see the extended language network (ELN), Ferstl and Cramon, 2008). Recently, a study on the interpretation of conversational implicatures has highlighted this region as sensitive to the level of implicitness of the speaker's message (Jang et al., 2013). On a lower-level of processing, ATLs are considered a domain-general semantic hub (e.g. Visser & Lambon Ralph, 2011) but a recent review by Wong and Gallate (Wong & Gallate, 2012) of literature on ATLs and processing of socially relevant stimuli suggests that this region could nevertheless be biased towards processing social information or personally relevant stimuli.

Middle and posterior temporal cortices are also part of the ELN, where they are ascribed a role of in integration and interpretation of language. Right middle temporal cortex is commonly seen in tasks going beyond literal meaning of utterances, including figurative language processing such as metaphors (Bottini et al., 1994), or idiomatic expressions (Lauro, Tettamanti, Cappa, & Papagno, 2008; Proverbio, Crotti, Zani, & Adorni, 2009; Zempleni, Haverkort, Renken, & A. Stowe, 2007). Findings by Kuperberg et al. (2000) and Kircher et al. (Kircher, Brammer, Andreu, Williams, & McGuire, 2001) indicate that right middle and superior temporal regions might be sensitive to the ease of semantic integration.

Taking a step back, we have seen that in the current job interview paradigm, the neural systems involved in cognitive perspective taking (mPFC, TPJ, precuneus) and in other cognitive operations associated with more complex discourse (IFG, ATL) are both responsive to face-saving indirectness, in a way that does not depend on whether the participant is being addressed or merely overhearing. However, as for core nodes in the *affective* ventral salience network, bilateral anterior insulae and ACC, things are a little different. For active addressees, we had not only predicted an indirectness effect in these areas, but also a *stronger* indirectness effect here in addressees than in active overhearers. After all, evasive or otherwise 'manipulative' indirect replies can be expected to elicit a stronger emotional response if *you* are the target of such social navigation than when the replies are directed at somebody else. Both predictions were clearly confirmed. As can be seen in Fig 3C, face-saving indirect replies reliably increased activity in the anterior insulae and ACC when participants were being addressed, and did so much more than when they were overhearing (no effect of indirectness).

In the light of the ROI findings in the more cognitive areas displayed in Figure 3A and 3B, the ventral salience network result displayed in Fig 3C is interesting, for various reasons. First, it suggests that indirectness does not additionally activate *every* area inspected in a powerful ROI analysis. It is also interesting because for participants, there was no *functional* difference in the task between the two conditions: in both situations, they really only needed to extract the relevant information, compare them to their expectations about a perfect candidate, and make an informed judgment. What our findings suggest, however, is that this is not the whole story: being addressed does change things, not so much because of changes at the referential level (what is being talked about), but because the *social* configuration is different. With the current paradigm lacking visual contact (i.e., lacking feedback from gestures or facial expressions) and true relevance for participants (such as bearing the consequences for a badly chosen employee), these social-affective differences between being addressed and overhearing are bound to be much larger in real life.

How crucial is the emotional involvement in interpreting indirect replies, observed in addressees, to the enterprise of understanding language comprehension? Is it a mere downstream affective consequence of “getting the message” – feeling in a particular way about how the candidate handled a sensitive question? Or is it an inextricable part of the interpretation process, a “hot” addition to the primarily “cold” analysis of speaker meaning? We have no means to disentangle these two positions in our study. However, the issue goes to the heart of where linguistic analysis stops and social cognition and affect begin. We suspect that although these domains may be separated analytically, they are in fact inextricably intertwined (see Van Berkum, 2015) – real language use is inevitably also social action, and language interpretation is therefore also deeply social and affective.

Finally, how "language-specific" are our findings on linguistically induced cognitive perspective taking and other cognitive processes associated with linguistic discourse? We think the most sensible position here is to accept that most of the indirectness effects we report here will *not* be language specific. For one, cognitive perspective taking is not limited to linguistic discourse, neither in pragmatic accounts of inferential communication (e.g. Levinson, 2006; Tomasello, 2008), nor in how the associated network can be made to light up in fMRI studies (e.g., see Mar, 2011, for a meta-analysis of verbal and non-verbal ToM studies). Furthermore, major nodes of the latter mentalizing network have also been associated with other functions (see, e.g., Binder & Desai, 2011; Cabeza, Ciaramelli, & Moscovitch, 2012; Humphreys & Lambon Ralph, 2014). Fronto-temporal areas that are more heavily taxed by the complexity inherent in face-saving indirect replies may involve working memory, episodic memory, and executive control over both (e.g., via selection or inhibition; see Fedorenko, 2014), again presumably generic processes, recruited in the service of language comprehension. And of course nobody would claim that salience-related affective processing is specific to language. Our prediction, therefore, is that a non-verbal design with face-saving indirect gestures should generate comparable results. On the whole, language processing seems to require a number of networks, including core networks for memory and unification, with the recruitment of additional networks that support the rich cognitive and affective infrastructure in which the language system is embedded (see Hagoort, 2005, 2013, 2014, and Van Berkum, 2015, for perspective and details).

5. Conclusion

We have seen that the comprehension of face-saving indirect replies, an important device that communicators use to navigate the social world, selectively recruits virtually all of the brain areas in active addressees and active overhearers that we also saw recruited in a "passive" experimental design, with passive overhearers. This is quite compelling, as the replication was carried out with a different task, different stimuli and different participants, and with different 'faces' to be saved, for different reasons. Important core processes of understanding what the speaker is *really* saying thus appear to be stable across a variety of settings, i.e. can be observed both in a passive listening task that is representative of the majority of neurobiology of language experiments, and in a more natural action oriented overhearing task where the focus was on "doing things with language".

In line with pragmatic analysis (e.g., Grice, 1975; Clark, 1996), understanding language heavily relies on recognizing the intention of the communicative partner. One of the original contributions of this study was to move the focus from examining language-as-product to language-as-action – that is, the participants' goal was to do something with the information derived from direct and indirect utterances, not just simply listen to them. We showed that the basic infrastructure for cognitive perspective taking and for dealing with the complexity of an extended discourse does not significantly change in such an active paradigm. We also showed that for these cognitive operations, it doesn't matter whether the face-saving indirect replies are directed towards the participants themselves, or to a third person -- presumably because the same inferential and representational work needs to be done in either situation. What *does* change with listener status is the activity in brain areas associated with emotional salience. This makes sense: whether somebody is navigating *you* or somebody else with evasive indirect language should matter *somewhere*, and affective responses to what a person is trying to achieve, i.e., his or her social move, is an obvious place to look.

References

Bambini, V., & Bara, B. G. (2012). Neuropragmatics. In J.-O. Östman & J. Verschueren (Eds.), *Handbook of Pragmatics*. Amsterdam/Philadelphia: John Benjamins.

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology*, *23*(3), 361–372.

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the language given: the neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578.

Bernhardt, B. C., & Singer, T. (2012). The Neural Basis of Empathy. *Annual Review of Neuroscience*, *35*(1), 1–23.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer. *Computer program version 4322*. www.praat.org.

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language-A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, *50*(11), 2669–2683.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., & Frith, D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language: A positron emission tomography activation study. *Brain*, *117*(6), 1241–1253.

Boyd, B. (2009). *On the Origin of Stories: Evolution, Cognition, and Fiction*. Harvard University Press.

Brett, M., Anton, J.-L. L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. In *NeuroImage* (Vol. 16, p. abstract 497).

Brown, P., & Levinson, S. C. (1987). *Politeness: some universals in language usage*. Cambridge University Press.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.

Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. *Communication Theory*, *18*(2), 255–280.

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends in Cognitive Sciences*.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, H. H. (2006). Social actions, social commitments. In Stephen C. Levinson & N. J. Enfield (Eds.), *Roots of human sociality: Culture, cognition, and human interaction.* (pp. 126–150). Oxford: Berg Press.

Cohen, J. (2001). Defining Identification: A Theoretical Look at the Identification of Audiences with Media Characters. *Mass Communication & Society*, *4*(3), 245–264.

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *The Scientific World Journal*, *6*, 1146–1163.

Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews*, *35*(3), 903–11.

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in psychology*, *5*, 335.

Ferstl, E. C., & Cramon, D. Y. Von. (2002). What Does the Frontomedian Cortex Contribute to Language Processing : Coherence or Theory of Mind ? *NeuroImage*, *1612*, 1599–1612.

Ferstl, E., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A metaanalysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581–593.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday Anchor.

Gottschall, J. (2012). *The storytelling animal: How stories make us human.* Houghton Mifflin Harcourt.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of personality and social psychology*, *79*(5), 701–721.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax And Semantics* (Vol. 3, pp. 41–58). Academic Press.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*.

Hagoort, P., Baggio, G., & Willems, R. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 819–836). Cambridge, MA: MIT Press.

Hagoort, P., Hald, L., Bastiaansen, M. C. M., & Petersson, K.-M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science (New York, N.Y.)*, *304*(5669), 438–41.

Hagoort, P., & Levinson, S. C. (2014). Neuropramatics. In Gazzaniga M.S. & G. R. Mangun (Eds.), *The cognitive neurosciences* (pp. 667–674). Cambridge, Mass: MIT Press.

Hervé, P.-Y., Razafimandimby, A., Jobard, G., & Tzourio-Mazoyer, N. (2013). A shared neural substrate for mentalizing and the affective component of sentence comprehension. *PloS one*, *8*(1), e54400.

Hoeks, J. C., & Brouwer, H. (2014). Electrophysiological Research on Conversation and Discourse. In T. M. H. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology* (pp. 365–386). New York: Oxford University Press.

Holtgraves, T. M. (2002). *Language as social action: Social psychology and language use.* Mahwah, NJ.: Erlbaum.

Humphreys, G. F., & Lambon Ralph, M. A. (2014). Fusion and Fission of Cognitive Functions in the Human Parietal Cortex. *Cerebral Cortex*.

Jang, G., Yoon, S. A., Lee, S. E., Park, H., Kim, J., Ko, J. H., & Park, H. J. (2013). Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. *NeuroImage*, *81*, 61–72.

Kircher, T. T. J., Brammer, M., Andreu, N. T., Williams, S. C. R., & McGuire, P. K. (2001). Engagement of right temporal cortex during processing of linguistic context. *Neuropsychologia*, *39*(8), 798–809.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: an fMRI study of causal inferencing across sentences. *NeuroImage*, *33*(1), 343–61.

Landauer, T., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2), 259–284.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Lauro, L. J. R., Tettamanti, M., Cappa, S. F., & Papagno, C. (2008). Idiom comprehension: a prefrontal task? *Cerebral cortex (New York, N.Y. : 1991)*, *18*(1), 162–70.

Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: the theory of the strategic speaker. *Psychological review*, *117*(3), 785–807.

Levinson, S. C. (2006). On the human "interaction engine." In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39–69). Oxford: Berg.

Li, S., Jiang, X., Yu, H., & Zhou, X. (2014). Cognitive empathy modulates the processing of pragmatic constraints during sentence comprehension. *Social cognitive and affective neuroscience*, *9*(8), 1166–74.

Mar, R. A., & Oatley, K. (2008). The Function of Fiction is the Abstraction and Simulation of Social Experience. *Perspectives on Psychological Science*, *3*(3), 173–192.

Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain." *Frontiers in Human Neuroscience*, *6*.

Mason, R. A., & Just, M. A. (2011). Differentiable cortical networks for inferences concerning people's intentions versus physical causality. *Human Brain Mapping*, *32*(2), 313–29.

Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. A. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. *NeuroImage*, *37*(3), 993–1004.

Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The neurocognition of referential ambiguity in language comprehension. *Linguistics and Language Compass*, *2*(4), 603–630.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in human neuroscience*, *3*(July), 14.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(3), 833–838.

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227–1235.

Proverbio, A. M., Crotti, N., Zani, A., & Adorni, R. (2009). The role of left and right hemispheres in the comprehension of idiomatic language: an electrical neuroimaging study. *BMC neuroscience*, *10*(1).

Rapp, A. M., Mutschler, D. E., & Erb, M. (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage*, *63*(1), 600–610.

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Consciousness and Cognition*, *17*(2), 457–467.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in human neuroscience*, *7*, 712.

Slater, M. D., & Rouner, D. (2002). Entertainment-Education and Elaboration Likelihood: Understanding the Processing of Narrative Persuasion. *Communication Theory*, *12*(2), 173–191.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39.

St George, M., Kutas, M., Martinez, A., & Sereno, M. I. (1999). Semantic integration in reading: Engagement of the right hemisphere during discourse processing. *Brain*, *122*(7), 1317–1325.

Stolk, A., Verhagen, L., Schoffelen, J.-M., Oostenveld, R., Blokpoel, M., Hagoort, P., … Toni, I. (2013). Neural mechanisms of communicative innovation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14574–9.

Taylor, K. S., Seminowicz, D. A., & Davis, K. D. (2009). Two systems of resting state connectivity between the insula and cingulate cortex. *Human Brain Mapping*, *30*(9), 2731–2745.

Tomasello, M. (2008). *Origins of Human Communication. Communication* (p. 379).

Tylén, K., Wallentin, M., & Roepstorff, A. (2009). Say it with flowers! An fMRI study of object mediated communication. *Brain and Language*, *108*(3), 159–166.

Van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S.-A. (2012). Pragmatics in Action: Indirect Requests Engage Theory of Mind Areas and the Cortical Motor Network. *Journal of Cognitive Neuroscience*, *24*(11), 2237–2247.

Van Berkum, J. J. A. (2009). The neuropragmatics of "simple" utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave Macmillan.

Van Berkum, J. J. A. (2010). The brain is a prediction machine that cares about good and bad—any implications for neuropragmatics. *Italian Journal of Linguistics*, *22*(1), 181–208.

Van Berkum, J. J. A. (2015). Language comprehension and emotion: where are the interfaces, and who cares? To appear in M. Miozzo, G. de Zubicaray, & N. O. Schiller (Eds.), *Oxford Handbook of Neurolinguistics*. Oxford: OUP.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*(2), 147–182.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*(3), 564–584.

Visser, M., & Lambon Ralph, M. A. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *Journal of cognitive neuroscience*, *23*(10), 3121–3131.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain*, *129*(4), 932–943.

Wong, C., & Gallate, J. (2012). The function of the anterior temporal lobe: A review of the empirical evidence. *Brain Research*, *1449*, 94–116.

Zempleni, M. Z., Haverkort, M., Renken, R., & A. Stowe, L. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *NeuroImage*, *34*(3), 1280–1291.

Zwaan, R. a, & Radvansky, G. a. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162–85.

Acknowledgments

We want to thank Laura Arendsen, Brenda Lelie, José Kivits and Marjolijn van Gelder for recording the questions and replies used in the Job interview, as well as Paul Gaalman for technical assistance with data acquisition. JB was partly supported by VEGA grant 2/0154/13, JvB was supported by NWO Vici grant #277-89-001.

Supplementary materials

1. Q: How good are your typing skills?
R: Quite normal, in my opinion. My previous employer was satisfied with me.

2. Q: Do you speak any foreign languages?
R: Uhm… I speak very good German but not any other languages. Well, I also speak some English but I am much better in passive English than in really… really using English.

3. *Q: How are you going to improve your language skills?*
R: I am planning to follow an English course this summer.

4. Q: Do you have any experience with writing yearly reports?
R: Hm, I don't really have any experience but I've read some on the internet. It does not look too difficult and I am a really good writer.

5. Q: Have you applied for any other positions before applying for this job?
R: Yes, I've applied in every bigger city in the country apart from the one where I live.

6. Q: You have a gap of several years on your CV since your last job**. Is this the first job interview you have been invited to?**
R: Finding a job in my area is extremely difficult.

7. Q: Do you have experience with business correspondence in English?
R: I remember that we had to write business letters in school, but not too often.

8. *Q: Why do you fit into this line of work – personality-wise?*
R: I have all the personal qualities to work with finances.

9. **Q: I can see that your previous job lasted less than three months. Did you resign voluntarily?**
R: I came there in the middle of a restructuring process and the newcomers had to leave as the first ones.

10. Q: Do you have any experience with electronic scheduling programs?
R: Eeh, yes, I have tried several different types and I can work with them very well.

11. *Q: Which of the skills you listed on your CV have you not yet had the chance to apply into practice?*

R: I have never needed to work with Powerpoint.

Example of one job interview. Direct question-reply pairs are in *italics*, indirect question-reply pairs are in **bold**. The rest are filler items. The interview is for the position of office manager. The interviews were conducted in Dutch.

SI: Simple main effects of indirectness within addressees or overhearers

Table S1: *Brain areas activated in the contrasts Indirect Addressee > Direct Addressee and Indirect Overhearer > Direct Overhearer. Clusters are shown at a threshold of 0.001 voxel level and extent of >200 voxels.*

Anatomical region	Coordinates of local maxima			Cluster size	p-value (cluster-level FWE corrected)
	x	y	z		
A) Indirect > Direct Addressee					
R middle temporal gyrus	48	-26	-6		
R middle temporal gyrus (temporal pole)	48	4	-30		
R inferior frontal gyrus (pars triangularis)	52	26	4		
	54	20	20		
R inferior frontal gyrus (pars orbitalis)	48	30	-6	5510	<0.001
R superior temporal gyrus	60	-16	0		
	60	-40	12		
R thalamus	10	-12	0		
	18	-16	0		

R insula	34	20	-8		
	34	22	-4		
R inferior temporal gyrus	50	10	-14		
R pallidum	26	-10	-2		
L middle temporal	-50	-36	0		
	-66	-30	0	1118	<0.001
L superior temporal	-44	-24	12		
L inferior frontal gyrus (p. triangularis)	-56	20	10		
L superior temporal pole	-38	14	-18		
	-44	26	-4	1221	<0.001
L inferior frontal gyrus (p. orbitalis)	-36	18	-16		
	-46	32	-4		
L superior frontal gyrus	-18	52	30		
	-4	50	30		
L supplementary motor area	-10	12	64		
	-6	14	58		
R superior medial frontal gyrus	2	56	24	1987	<0.001
L anterior cingulate cortex	-8	40	24		
L superior medial frontal gyrus	-6	52	14		
	6	44	46		
R superior frontal gyrus	10	42	38		
	14	42	48		
L angular gyrus (TPJ)	-54	-56	22		
L superior temporal gyrus	-64	-52	16	244	0.098
L middle temporal gyrus	-58	-56	28		

SI: Main effects for being addressed versus overhearing -- Job Interview task

A) Indirect > Direct Overhearer

R middle temporal gyrus	52	-28	-6	1655	<0.001
	58	-8	-20		
R angular gyrus (TPJ)	60	-56	32	631	0.004
R medial temporal pole	48	6	-34	233	0.110
R inferior temporal gyrus	50	-6	-28		
R inferior frontal gyrus (p. triangularis)	50	28	4		
	52	24	8		
R inferior frontal gyrus (p. opercularis)	48	18	36	1109	<0.001
	42	14	36		
R middle frontal gyrus	44	10	48		
	44	16	50		
R superior medial gyrus	2	56	20		
	4	40	46		
	12	62	20		
L superior medial gyrus	-4	48	38	704	0.002
	2	36	50		
	-8	34	56		
	-6	56	30		
L inferior frontal (p. triangularis)	-54	22	8	358	0.033
	-46	26	-2		
	-46	24	6		
	-48	24	2		
L inferior frontal (p. orbitalis)	-46	32	-4		
L middle temporal gyrus	-54	-26	-4	291	0.062
	-48	-36	0		

The comparison of the addressee condition to the overhearer condition (*Addressee > Overhearer*), pooled across direct and indirect replies, yielded one cluster in the right hippocampus and fusiform gyrus (see Figure S2), with a small portion of the cluster extending into the right amygdala.

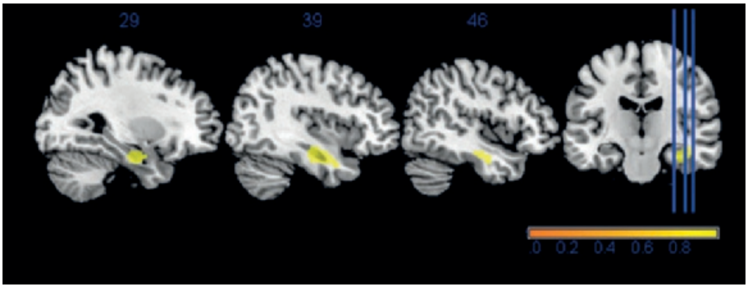


Figure S2: Significant activations for the Addressee vs Overhearer comparison. Significant effects are displayed on sagittal slices (x coordinate levels are in millimeters: 29, 39, 46). Color bar denotes t-values (in blue).

The opposite comparison (*Overhearer > Addressee*) involved an extensive cluster of cortical and subcortical regions, dominated by bilateral pre- and postcentral gyri and bilateral cerebellum. In addition, there were bilateral temporal and inferior frontal activations, parietal cortex, cuneus, and thalamic activations, as well as activations in anterior and middle cingulate cortex (see Figure S3).

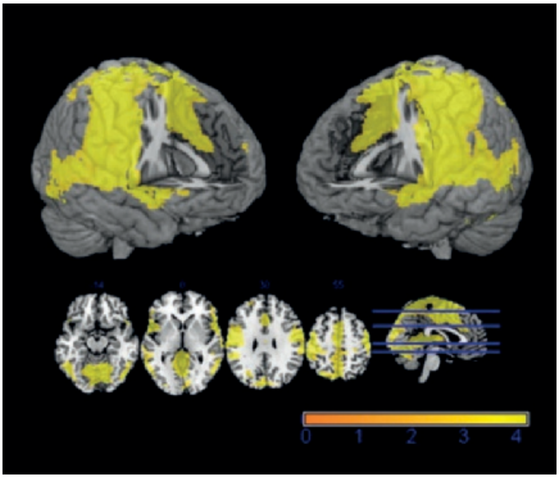


Figure S3. Significantly activated cluster in the Overhearer > Addressee contrast. Color bar denotes t-values (in blue).

In the regions-of-interest analyses, we obtained a main effect of listener status for mentalizing areas, with stronger activation in each ROI for both direct and indirect replies in the *overhearer* condition than in the *addressee* condition ($F(1,19) = 7.79$, $p = 0.012$). There was also a significant main effect of listener status in the regions-of-interest analysis on discourse-level cognitive processes ($F(1,19)= 9.92$, $p = 0.005$), with again more activation to direct and indirect replies when participants were overhearing them than when they were addressees. In the affective salience network analysis, no significant main effect of listener status was observed ($F(1,19)=2.77$, $p=0.112$).

SI: Main effects for being addressed versus overhearing -- Control task

We also examined the difference between being addressed or overhearing in a minimal, essentially non-communicative control utterances comparison, outside of the Job Interview task. In the Control Addressee condition, participants were told "listen to the candidates reading these sentences for you", and in the Control Overhearer condition, they were told to "listen to these sentences recorded at a previous session with the candidates". The comparison of these isolated control utterances (*Control Addressee > Control Overhearer and vice versa*) yielded no significantly activated brain areas, suggesting that the above

Addressee vs Overhearer main effects in the Job Interview task were specific to the communicative setting of that task.

Discussion of main effects of listener status

In line with the aims of our study, our predictions were focused on the *differential impact* of face-saving indirectness, and its modulation by listener status, in the Job Interview task. We did not have strong a priori hypotheses for the whole-brain comparison of addressee and overhearer conditions. One of the most demanding aspects of the Job Interview task, as conveyed by the participants in the exit interview, was to keep track of whether the candidate fulfils the job requirements stated in the job ad before each round of interviews. At the same time, the participants kept updating their impressions about the candidates based on their replies, forming positive or negative impression as they progressed through the interview. A common comment in the exit interview was that it was more difficult for participants to remember the incoming information and make judgment in a relatively fast-paced process when they also had to concentrate on leading the interview themselves (in the addressee condition). In view of such post-experiment comments, one might expect that the addressee condition would have more activity in areas implied in working memory, cognitive flexibility or cognitive control.

As can be seen in Figure S2, the Addressee > Overhearer contrast did not support this idea; only a small right-hemisphere medial temporal lobe cluster emerged, whereas the reverse Overhearer > Addressee contrast yielded a much larger cluster of differential activations. Importantly, the critical whole brain analysis activations for indirectness reported in Figures 1 and 2 in the main paper were in *different* regions of the brain, revealing that, whatever the cause of these overall listener status effects in the Job Interview task, they are effectively partialled out in our critical *indirect versus direct reply* contrast. As for the (non-predicted) overhearing effect in mentalizing and fronto-temporal ROIs, we suspect it reveals that overhearers are also taking the perspective of the *other interview*er into account. Finally, the absence of significantly activated regions in the control contrasts (Control Addressee > Control Overhearer or vice versa) suggests that the Listener status effect in the job interview setting is not simply due to being directly spoken to or not.

CHAPTER 4: GESTURAL INDIRECTNESS: A BEHAVIORAL STUDY

Did you get my point? How recipients understand indirect communicative gestures.

Jana Bašnáková[1,2], Jos van Berkum[3], Emanuela Campisi[4], Andrej Mentel[5], Peter Hagoort[1]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands [2] Institute of Experimental Psychology, CSPS SAS, Slovakia [3] UiL-OTS, Department of Languages, Literature and Communication, Utrecht University, Netherlands [4] Department of Humanities, University of Catania, Italy [5] Institute of Social Anthropology, Faculty of Social and Economic Sciences, Comenius University, Slovakia

Abstract

Communication is not limited to the verbal domain, and communicative signals can be a lot more than just conventional and coded. By exploiting various levels of contextual information, recipients can express and understand meanings conveyed by gestures, facial expressions, winks, or even well-timed silence. Our reaction time study compared the time-course of communicative meaning comprehension in cartoon stimuli in two different modalities: verbal utterances and pointing/showing gestures, which were either simple (direct) or complex (indirect). The choice of indirect speech and pointing gestures was motivated by the fact that they *amplify* two features present in everyday communication: first, the complete reliance on context for figuring out the speaker's communicative intentions in case of pointing gestures lacking any coded meaning; second, the necessity to bridge the gap between explicitly stated and implicitly communicated message, which is large with indirect speech. We showed that, even though indirect verbal replies and pointing gestures ultimately lead to the same mental representations and were motivated by identical social intentions, indirect gestures took much longer to interpret than their verbal counterparts. The locus of this difference was not just general difficulty interpreting gestures, as there was no corresponding difference between the two direct replies. Rather, as suggested by the pattern of reaction time results and cognitive empathy measures (IRI), when context gets less constraining, the focus of indirect gestures – their literal meaning - becomes more ambiguous to interpret.

Based on:

Jana Bašnáková, Jos van Berkum, Emanuela Campisi, Andrej Mentel & Peter Hagoort (submitted). Did you get my point? How recipients understand indirect communicative gestures.

1. Introduction

Communication is not limited to the verbal domain, as people are able to skilfully express and understand meanings conveyed by gestures, facial expressions, or even slight movements such as winks. Simple communicative gestures like pointing are often exploited, and easily understood, for example when the environment is too noisy, or when communication needs to be hidden from somebody else's attention. Parents of small children are probably familiar with situations when they try to inconspicuously point to things for each other and pantomime behind the children's backs to agree on an efficient course of action, knowing that saying aloud taboo words such as "candy" or "ice-cream" might trigger a difficult situation.

Although most psycholinguistic studies are focused on characterizing communication via the linguistic code, a communicative signal can be a lot more than just conventional and coded (Clark, 1996). Our ability to understand signals such as gestures, even without any verbal content, is likely subserved by much of the same social-cognitive infrastructure necessary to infer the speaker's communicative intentions from speech. Clark has framed this in terms of various *methods of signalling*. Different communicative devices –sentences, gestures – deliberately performed for the listener serve as signals informing the recipient about the speaker's goal. When someone is asked whether there is a nice movie in the cinema tonight, the utterance "*I don't know*," or a shoulder shrug are both equally good ways of answering, and both ultimately result in the same mental representation of the speaker's intent in the mind of the recipient. When a two-year old wants to be picked up by his mom, he can say "*Up!*" or simply extend his arms towards her with an expectant look. Linguistic and non-linguistic signals can be combined into composite signals (Clark, ibid., Enfield, 2009; Kendon, 2004), and in real-life interactions are often presented in multiple channels simultaneously.

Pragmatic theories single out two more ingredients which underlie communication, apart from the overt signal. First, the interpretation of signals relies on interlocutors having established common ground (Brennan & Clark, 1996; Karttunen & Peters, 1975). That is, listeners interpret the signals they see or hear with reference to beliefs they share with the speaker, either by virtue of physical co-presence, or shared cultural beliefs, or what was said in previous conversation (Clark, 1996, Tomasello, 2008). A necessary condition of common ground is its mutuality – not only the interlocutors having similar beliefs and knowledge, but also knowing that they *both* have them. To paraphrase Tomasello's example, if two friends are walking towards the library and one points to a bicycle standing in front of it, the other must draw on beliefs they mutually share to infer what the gesture was supposed to mean. If this is the bicycle of the recipient's ex-boyfriend, and the recipient knows that the communicator knows about the existence of such a bike, as well as about the break-up, then the pointing gesture can be interpreted as "*Let's not go inside, your ex is there!*". On the contrary, if the recipient knows that the communicator has no knowledge about the relationship's unhappy ending, then her gesture would probably be taken to mean "*Look, your boyfriend's here, let's go in!*". Or, if the communicator knows nothing about the relationship at all, but is a bike enthusiast, then the addressee takes the gesture to express just a simple "*Look, what a cool bike!*"

The second ingredient in our "communication toolkit" is inferencing. Overt signals, such as gestures or sentences, often underestimate what the speakers really want to communicate, and thus can be rather viewed as a starting point. If listeners want to get from this starting point to the "finish line" of fully interpreting the speaker's message, they must make inferences. There are many different types or levels of inferences, from causal inferences connecting different events within and across utterances, to mentalizing inferences when one wants to get at parts of the messages which were intentionally left implicit (Harnish & Garret, 2007). In our previous work, we have targeted pragmatic inferences necessary to interpret indirect replies with face-saving motives, such as not wanting to hurt the feelings of the listener (Bašnáková, van Berkum, Weber, & Hagoort, 2015; Bašnáková, Weber, Petersson, van Berkum, & Hagoort, 2014). However, there is evidence that mentalizing inferences are necessary to interpret other pragmatic phenomena as well (e.g. Garrett & Harnish, 2007), even simple direct single-word speech acts (Egorova, Shtyrov, & Pulvermüller, 2016). Inferencing and common ground are often related, as inferences rely on common ground, and interlocutors must also frequently "work out" common ground, unless it is immediately obvious to both communicative partners.

This study aims to explore the time-course of speaker meaning interpretation via two relatively little researched areas which crucially depend on both exploiting common ground between the speaker and the listener, and on making pragmatic inferences. First, we investigated the processing of communicative pointing gestures as an instance of signals that

completely lack any intrinsic/coded meaning (Cartmill, Beilock, & Goldin-Meadow, 2011), and therefore fully depend on context for interpretation. Second, we focused on indirect speech acts whose interpretation requires bridging the gap between their explicit and implicit meaning via pragmatic inferences. We explore pointing and indirectness together in a single study because, as will become clear later, they may interact in interesting ways.
We created a cartoon paradigm (*Figure 1*) which requires the participants to interpret direct and indirect replies communicated via a verbal utterance or a pointing/showing gesture, and measured how fast they can process the resulting four types of communicative acts. In the rest of the Introduction, we will summarize empirical research on indirectness and communicative pointing, and present our study design and predictions.

1.1. Indirectness in communication

Being indirect is a communicative strategy often used as a way to save one's face, or to be polite in conversations (Brown & Levinson, 1987; Holtgraves, 1998), and most of our previous research explored this type of indirectness. However, there are also other motives for speaking indirectly, such as plausible deniability (Pinker, Nowak, & Lee, 2008), and "information economy", the efficient packaging of information into a single utterance. As for the latter, when a speaker wants to reply not just to the question asked but to helpfully pre-empt further questioning, she can address what she thought was the *motivation* behind the question in the first place (Walker, Drew, & Local, 2011). For example, when one housemate asks the other whether he wants to join him for a run, then the housemate can answer indirectly, detailing the reasons why he cannot go ("*My running shoe is worn out,*"). He could choose a more direct reply "*No, I will not,*" but it would be much less informative, leading to further clarifying questions. In situations like these, indirect answers require the listener to do some inferencing in order to get from the literal meaning of the utterance (*his shoe is worn out*) to its implicit content (*No, he will not join me*). At the same time, however, indirectness will save him the effort necessary to follow up the original question to get at the information that he needs.

There is a large body of work on "special" non-explicit meaning comprehension, such as irony, sarcasm or metaphors, within experimental psychology and pragmatics. However, regarding "regular" indirect speech acts in interpersonal communication, only a few authors propose processing models of the different kinds of inferences and cognitive processes involved (most notably, Carston, 2013; Recanati, 2004; Sperber & Wilson, 1995; Wilson & Sperber, 2004).

1.2. Pointing gestures and pragmatics

Interpreting signals about communicative intentions via various channels, including nonverbal ones, is at the heart of communication (Kita, 2003; Schulze, Grassmann, & Tomasello, 2013). Even though gestures are rarely explored from a pragmatic perspective, there are exceptions. Kendon (2004) has introduced the term "pragmatic function of gesture" to describe meanings (co-speech) gesture obtain when they are used as performatives (see also Enfield, Kita, & de Ruiter, 2007). Tomasello's work (Tomasello, Carpenter, & Liszkowski, 2007) has highlighted what underlies communicative meaning interpretation using the example of simple pointing gestures. Unlike emblems (*thumbs up/waving goodbye*), pointing does not carry any referential meaning in itself, but can still express a variety of meanings depending on how it is used in context. Interpreting declarative pointing, i.e. pointing to share attention to an outside entity, presupposes that the communicator and the recipient both share a common ground or joint attentional frame, that the recipient recognizes the communicator's social intention and recovers the goal behind her pointing action. This rich view on declarative pointing has been supported in a neuroimaging study which showed that declarative, but not imperative, pointing engages brain regions associated with mentalizing (Brunetti et al., 2014).

1.3. Pointing indirectly

The only two studies so far combining gestural communication and indirectness from the listener's perspective come from developmental psychology[6]. Schulze and Tomasello (2015) have shown that infants as young as 18 months are able to understand indirect gestural communicative acts, as long as these nonverbal signals are ostensive. The authors make an explicit comparison between gestural and verbal communication, stating that the intentional-inferential structure of human communication is the same in both of these domains (see also Tomasello, 2008). From this perspective, when a child is looking for a missing piece of a toy

[6] For completeness, Kelly (2001; Kelly, Barr, Church, & Lynch, 1999) also studied gestures and indirect speech acts, but his focus was on the facilitatory effects gestures have on interpreting indirectness, rather than on the cognitive processes of gesture interpretation as such.

in the presence of a locked container, and the adult either says "I have the key", or ostensively points to the key, it will be equally difficult for the child to infer what the adult is implying -- that the toy is in the container.

Another study comparing indirect speech acts in both verbal and gestural domain (Bucciarelli, Colle, & Bara, 2003, see also Cutica, Bucciarelli, & Bara, 2006 and Cutica, 2005 for adult patients with LH- and RH-lesions) tested children between 2.6 and 7 years of age. Apart from indirect speech acts, they included a number of different types of pragmatic phenomena conveyed through verbal and gestural means, e.g. ironies or simple deceits. The indirect speech acts were both conventional (*pointing to the window that the actor wants to be closed*) or complex (*a girl pointing to a doll in the shop window and her sister showing her an empty purse indicating that she has no money to buy it)*. Participants made more errors interpreting complex indirect speech acts than conventional ones. However, it did not matter whether these acts were delivered via speech or gesture.

To explain these results, Bucciarelli et al. proposed a framework with two main explanatory factors: a) mental representations of varying complexity, and b) inferential load. Increasing complexity of representations and inferences presumably made it more difficult for children to understand a given speech act. There was, however, no difference predicted for the different modalities, since representations and inferential load should be, in principle, equally difficult regardless of whether they are delivered via speech or gesture.
Even though these two developmental studies give interesting insights into indirect speech acts processing in the gestural domain, the model they propose does not fully account for important differences of speaker meaning interpretation between the verbal and gestural domains. Below, we propose a more fine-grained model to characterize the cognitive architecture behind verbal and non-verbal indirectness, and ultimately, communicative meaning interpretation. Before describing it in more detail, we first present our experimental paradigm.

1.4. The cartoon paradigm

We conducted a study where participants had to interpret indirect and direct (baseline) speech acts delivered via gestural or verbal means, performed by characters in cartoons. Cartoons were chosen because they represent a genre that can portray real-life interactions in a simplified way and they work well with both speech and gestures.

Each cartoon depicted an interaction between two people, enacted over three frames. The first one introduced the story setting and the characters. For example, two male students standing together in a hallway, one dressed in sports gear, talking about how he is ready to go out for a run. The second frame continued in developing the interaction by introducing a question from one of the character, for example – "*Would you like to join me?*", or, "*Why can't you go for a run?*". The reply from the other character was delivered on the third, target picture, either via a gesture or an utterance (similar to a speech bubble). The speaker in the target frame could reply in one of four ways, corresponding to four experimental conditions: following an open-ended Wh- question asking for information, he would answer a) directly verbally "*My shoe is worn out*," or b) directly gesturally, by pointing to his torn shoe that he was holding in his other hand. Or, following a yes/no question "*Would you like to join me?*", he would answer c) indirectly and verbally "*My shoe is worn out*," or d) indirectly and gesturally by pointing to a torn shoe that he was holding in his other hand (see *Table 1* and *Figure 1,* also *Table 3* for additional examples). Importantly, the content – or the "utterance meaning" - of the target reply was identical for the two verbal and two nonverbal conditions, and as closely matched between modalities as possible. The form (gesture or text) differed between verbal and gestural conditions.

	Verbal	Gestural
Direct (*Why can't you go for a run?*)	*My shoe is worn out.*	Holds up a worn out shoe and points to a hole in the sole.
Indirect (*Would you like to join me for a run?*)	*My shoe is worn out.*	Holds up a worn out shoe and points to a hole in the sole.

Table 1. Examples of the four experimental conditions

Figure 1: Examples of direct and indirect verbal and gestural stimuli. Captions are approximate translations, the original items were in Dutch.

The role of the participants was to read through the cartoons and press a button as soon as they were able to fully understand the reply, which was our measure of the time it took to process its meaning. We presented a simple probe question about a possible continuation of the situation after each cartoon (*The boys are going for a run together/Only one of the boys will go out for a run*), to be able to tell whether participants correctly interpreted the replies.

1.5 How is verbal and gestural indirectness different? A processing model

Verbal and gestural indirect speech acts ultimately lead to the same mental representation/referential intention (*I can't go for a run with you right now*) and are motivated by the same social intention (*to inform friends about events of potential relevance*). On the surface, it might seem that the difference between them is just superficial, i.e. their different modalities. After all, regardless of modality, communicators in both cases do the same thing: they use signals to highlight a specific aspect of reality for recipients and expect, within the context of cooperative social exchanges, that she will take this isolated aspect of reality as maximally relevant in the given context and infer the message that they want to convey. For this reason, we agree with both Schulze and Tomasello (2015) and Buciarelli et al. (2003) that comprehension of gestural and verbal indirectness will most likely be supported by the same inferential-intentional cognitive architecture. However, based on a theoretical analysis presented below, we believe that there are aspects of interpretation of indirect verbal utterances and pointing gestures which nevertheless differ. Consequently, we predict that gestural communication used in more complex situations taxes some of the underlying meaning-making processes to a larger extent than its verbal equivalent.

In pragmatics, meaning construction is traditionally characterized as a two-tier process (e.g. Grice, 1975). First, there is *utterance-level* meaning, which describes the meaning of words and syntactic rules connecting them, uttered within a specific context. For example, pronouns are disambiguated and word meanings are selected, such that "I" refers to a particular person and "library" refers to a particular building on campus and not, for example, a specific bookshelf in one's home. This utterance-level meaning is routinely described as "what is said" (Gibbs, 2002). Next, there is speaker-meaning, or "what is implicated". This is the message that the speaker actually wants to communicate (i.e. not simply "*My shoe is worn out*", but rather "*I can't go running with a shoe like this one!*") and is often very different to the utterance's literal content. For completeness, various pragmatic theories place the boundary between what is said and what is implicated in different places, and there is also disagreement about the nature of the different inferential processes involved (see e.g. Garrett & Harnish, 2007 or Kissine, 2015). Despite discussion about the details, however, pragmatic theory suggests that context plays a role in utterance interpretation at various stages and in various ways, and even seemingly "ordinary" or "literal" sentences require contextual anchoring and enrichment.

What makes pointing gestures and indirectness interesting to study within a single paradigm is that, together, they pose a specific challenge at *each* of these two levels of meaning. Indirectness requires that listeners make mentalizing inferences about the speaker's intentions to account for a gap between their explicit and implicit meaning. At the same time, it is presumably harder for listeners to establish utterance meaning of nonverbal pointing signals in comparison to verbal signals, as they lack any coded meaning and must be entirely worked out from context. Crucially, pragmatic analysis predicts that making sense of an utterance at these two levels will not be independent, but that these processes will *interact*. As Clark (1997) points out, disambiguating what is said is often informed by what is implicated. This goes against the apparent equivalence of verbal and gestural indirect speech acts; it means that utterance-level meaning of indirect gestural signals is going to be more difficult to establish than utterance-level meaning of indirect verbal signals. In other words, listeners will have to work harder to understand an indirect gesture than an indirect verbal reply, even if the conversation and question preceding them will be identical.

Why? When it comes to simple mappings between what we signal and what we want our communicative partner to infer, such as in direct speech acts, words might not have any advantage over gestures. Both manage to highlight the relevant aspect of reality efficiently enough. But when it comes to conveying meanings which necessitate a longer inferential chain, such as indirect speech acts, verbal utterances have a clear advantage over gestures in that *they reduce the ambiguity inherent in the signal itself*. That is, when we are looking for evidence that directly prevents someone from running (direct utterance) versus when we are looking for *any* evidence informing someone's decision about going/not going for a run (indirect utterance), the question will constrain our search for what constitutes a suitable utterance-level meaning. In the indirect case, we will need to consider a larger interpretative space even if the reply – the pointing gesture – is physically identical. In theory, the same is true for both indirect replies, but since words are more efficient in minimizing ambiguity and accurately describing complex aspects of reality than gestures, we expect this process to be much less difficult in the indirect verbal case than in the indirect nonverbal case.

With this in mind, let us look in more detail at the exchange in the model item example. Two friends, A and B, are talking about going for a run. In the two direct conditions, it is already mutually known on the first context frame that B cannot join in. The target direct reply in both modalities (*My running shoe is worn out./Pointing to a worn-out shoe)* comes after a Wh- question (*Why can't you go running?).*

What is required in order to understand the intended meaning behind the direct reply? The piece of information indicated by the communicator B for recipient A is both simple and has been made very salient by the entire conversational context. By the time-point in the conversation when the verbal/gestural signal is performed, they both know what A is looking for. The question elicits a search for a *reason* preventing a healthy young man from going for a run; and a torn shoe, highlighted by a verbal utterance or by a pointing gesture, constitutes an easily accessible and relevant reason. In both cases, there will be some inferencing necessary to connect the question to the answer, but both the inferencing and the highlighting parts are likely to be comparably easy in the direct verbal and nonverbal conditions. It is even conceivable that, in the case of highly salient objects, pointing might actually be faster in highlighting the relevant object or situation in comparison to a verbal utterance because it circumvents linguistic processing.

In the two indirect conditions, the first context frame only offers information that A is getting ready for a run. The second frame contains an invitation – *Do you want to join me?* An invitation bids for acceptance or decline, and what follows is indirect decline: instead of just saying *no*, the communicator, B, gives a reason for why he has to decline (*My running shoe is worn out./Pointing to a worn-out shoe).* Here, as the communicated meaning becomes more complex, gestures present an additional challenge. In the indirect conditions, the relationship between the question and reply is less straightforward than in the previous case, where A knew he would be simply looking for evidence why his friend is not joining, and that anything highlighted either gesturally or verbally will count as such. Now, the *yes/no* question will leave A expecting a *yes/no* answer. However, what follows is not a clear yes or no, but only the starting point for inferring the "real" answer (in other words, *evidence towards* yes or no).

And here is where language makes things easier. Language gives us the tools to efficiently direct the listener's attention towards a specific aspect of reality. Therefore, when the answer to "*Will you join me [for a run]?*" is "*My running shoe is worn out.*", this has already made the inferencing less taxing because the speaker verbally highlighted what he considered the most relevant aspect of the present situation for the listener. In contrast, if such question is followed by a pointing gesture, we first have to establish the relevance of the highlighted part of reality in relation to the yes/no question (i.e., we have to determine “what is said”). Thus, while in verbal indirectness the starting point is already largely “assembled” by the speaker and we have to infer whether the answer is affirmative or negative, in gestural indirectness we are one step further away from getting at the intended meaning. Before establishing whether a signal constitutes evidence towards yes or no (a decision common to both indirect conditions), we first have to infer how the shoe is relevant to the invitation, i.e. what aspect of the shoe is my friend highlighting for me. Is it the existence of the shoe, is it the hole, is the speaker actually showing me his own shoe? This is reminiscent of the ‘indeterminacy of reference” argument made by Quine (1960), which is often used in language-acquisition literature – if we hear a word in an unknown language, how can we be sure what aspect of the situation this word refers to? Because, unless conditions are perfect and the gestural signal happens to highlights the relevant piece of reality flawlessly, a pointing/showing gesture in response to a question which does not very clearly specify the acceptable answer, will always be more ambiguous than a verbal utterance. The reason is that in verbal indirectness, all of the necessary decisions have already been made for the listener, since the speaker is presumably delivering an optimally relevant verbal message (“*My shoe is worn out*”). In other words, indirectness places additional constraints on how easily one can identify “what is said”, and since a pointing gesture is typically a much less clearly developed and formulated signal than a well-formed verbal utterance, the listener might find indirect gestures inherently more ambiguous.

To summarize, we agree with both Schulze and Tomasello (2015) and Bucciarelli et al. (2003) that the intentional-inferential architecture underlying indirectness in either modality is the same. However, our model proposes a difference on another level of meaning interpretation, i.e. in identifying “what is said”, or the gesture’s focus, which we believe is more difficult in the case of gestural indirectness.

1.6 Reaction-time predictions

Based on this reasoning, predictions for processing time differences among these four types of replies are as follows:

1. The two direct answers, verbal and gestural, should be interpreted equally fast, since they both involve two similar steps: First, identification of the message, i.e. recognizing the bike as ex-boyfriend's bike or the running shoe as worn out, or parsing the utterance "*My running shoe is worn out*"; secondly, inferring whether this counts as a relevant reason not to go running. Alternatively, if recognizing a visual scene is inherently simpler than processing linguistic code, gestural direct reply might be processed even faster than verbal direct reply.
2. The two indirect replies should be interpreted slower than their respective direct replies, since they involve at least one extra inferential step. The listener has to find out how the signal, equivalent to the explicit meaning of the utterance, is relevant to the yes/no question and infer whether the speaker wants to communicate a yes or a no (implicit part of the utterance).
3. Finally, the gestural indirect reply will take longer than the verbal indirect reply, because the recipient has to work harder to identify the *focus* of the pointing gesture. In the verbal indirect reply, this inference has been already made for the recipient. The communicator has planned her utterance in such a way as to be maximally efficient in conveying her message, and therefore reduced the inherent ambiguity in the signal.

1.7 Measures of individual differences

We also measured individual differences in participants' mentalizing abilities, in an attempt to relate them to individual differences in the processing of utterances and gestures. Our previous neuroimaging research on verbal indirectness has shown that mentalizing, or cognitive empathy, plays a role in interpreting indirect replies, since listeners need to get at the communicative intentions of speakers (Bašnáková et al. 2014, 2015). It is therefore likely that individual differences in the ability or propensity to infer the other's intentions will influence the speed of resolving indirect replies.

We used the Interpersonal Reactivity index developed by Davis (1983) which takes a multidimensional approach to measuring empathy. Its 28 items are divided into four subscales, and each subscale assesses a different aspect of empathy. Perspective Taking (PT) expresses the tendency to spontaneously adopt a psychological perspective of others, Fantasy scale (FS) measures the propensity to imaginatively transpose oneself into the feelings and actions of characters in fiction (books, movies, etc.), and Emphatic Concern (EC) evaluates differences in the extent to what people experience other-oriented feelings towards unfortunate others. Lastly, Personal Distress (PD) assesses self-oriented feelings that arise when a person sees others in unfortunate situations, such as personal unease or anxiety. In contrast to our previous studies, the present items do not specifically cover face-saving indirectness connected to emotional processing, and therefore we do not expect involvement of emphatic concern. However, we do expect that there will be differences in the speed of processing verbal and gestural indirectness for participants with low and high cognitive empathic abilities, expressed by the Perspective taking scale. We expect that the ability to adopt the psychological perspective of the other might be helpful in interpreting indirectness as such, since it requires making hypotheses about why the speaker uttered a seemingly irrelevant remark and what her communicative intentions were.

The Fantasy scale also seems to be relevant to gesture interpretation, and we included it for exploratory reasons. For example, participants with high level of imagination might find it easier to generate hypotheses about the meaning of gestures than participants with low level of Fantasy.

2. Materials and methods

We constructed 400 cartoon strips consisting of three pictures. There were always two “context” pictures and one “target” picture. On the first context picture, two characters would talk about something, and on the second one, one of the characters asked a question. Then, on the third picture, the second character gave a reply to this question, which served as the target reply. While the context pictures were always the same with regard to the drawing, there were two versions in terms of the accompanying dialog: a direct one and an indirect one. In the direct one, the question was always a “Wh-“ or “How” question. For example, “Why don’t you want to go for a run?”. In the indirect condition, the question was a yes/no question, for example “Do you want to join me?”.
The reply, given on the target picture, had four different versions, corresponding to the four experimental conditions. The two verbal conditions (direct and indirect) would show the

character's face and part of their body with text shown in the upper third of the picture. In terms of content, the replies in the two verbal conditions were identical (e.g. My running shoe is broken), but the preceding context rendered them either direct or indirect. The two nonverbal target pictures (again, direct and indirect) would show the participant pointing to some relevant situation/object, or making a "showing" gesture (i.e. holding up a worn-out running shoe), or a combination of those two (pointing and showing, e.g. pointing to a worn-out shoe which he was holding up for the other to see). We chose a pointing or a showing gesture purely on the basis of what was more natural and clearer for a given picture. Again, what would make the gesture direct or indirect was the preceding context including the question. Just like in the verbal versions, the gestures in the two gestural conditions were identical, corresponding to the content expressed verbally in the two verbal conditions. However, in some instances there were also subtle differences in the facial expression of the character when gestural reply was direct or indirect, to make them look more natural. Facial expressions accompanying gestures mark the social intention of the speaker (Clark, 1996), e.g. that it is an offer, or a decline.

Even though the content of the text in the direct and indirect *context* cartoons was slightly different, we equated it on overall length (in the number of words and syllables), as well as in terms of the overall semantic similarity of the context text to the verbal target reply. In some instances, the first context frames were reversed from left to right, because we needed to change the order of speakers.

2.1 Construction of the stimuli

In order to check whether our population (young native Dutch speakers) understands the cartoons in the same way as we intended them to, we conducted two pre-tests. The first one concerned the gestural versions only. In a paper-and-pencil test, we showed 20 participants the cartoons with the gestural replies, and asked them to interpret the target picture by writing down what would the person giving the reply say if they were to use words instead of pointing/showing gestures. Each participant saw only one version of each stimulus, either direct or indirect, i.e. we obtained 10 ratings for each version of a given item. We then counted how many replies accurately characterized the communicative intention of the speaker. For example, in the running-shoe example, an acceptable answer would be "*My shoe has a hole in it*," or "*I don't have a proper running shoe*," or "*I cannot run in this shoe*". An inacceptable answer would be "*I have new shoes*," or "*You can run in my shoes*". If a given cartoon received less than 8 acceptable ratings, we changed it.

The second pre-test concerned the perceived "directness" or "indirectness" of a reply. Again, in a paper-and-pencil version of the test, we presented 116 participants with our stimuli, and asked them to rate, on a scale of 1 to 7, how direct/indirect was the reply. The verbal and gestural versions were kept separate because we were primarily interested in the relative differences within each modality, and comparing the two modalities within one list might inflate the perceived differences between modalities. Just as in the first pre-test, no participant saw more than one version of each item. Even though the difference between direct and indirect replies within each modality was highly significant, in order to maximize the differences, we changed any replies in which the difference was less than 2 points, and re-run the changed stimuli with another 20 participants.

In the final stimulus set, the direct verbal replies were judged on average as 1.97 (0.69), direct gestural as 2.51 (1.02), indirect verbal as 4.83 (0.77) and indirect gestural as 4.23 (0.89). As expected, the differences between direct and indirect replies within each modality were highly significant (Wilcoxon sign-rank test, verbal $Z=-8.69$, $p<.001$, gestural $Z=-8.30$, $p<.001$).Unexpectedly, there was an interesting modality difference, as raters perceived the verbal items further apart in terms of indirectness than the gestural items. Direct gestural replies were rated as less direct than direct verbal replies, and indirect gestural replies were rated as less indirect than indirect verbal replies. We will return to this issue in the Discussion.

We constructed additional 25 filler items. The fillers were different than experimental items in that they would give an affirmative answer after an indirect question (while most experimental items gave a negative answers, e.g. that a common run is impossible), or that they would give an indirect answer after a wh- question.

In order to check whether participants understand the replies, we also constructed a series of probe sentences presented after each target reply which would describe a possible continuation of the interaction. Participants had to rate whether this outcome was "normal" or "unusual". For the worn-out shoe example, a *normal outcome* probe sentence would be "*Only one of the boys will go for a run*", whereas the *unusual outcome* sentence was "*Both boys are heading out for a run*". We equated the normal and unusual versions on the number of words. Each participant saw only one version of the probe sentence after a given item in each condition. We excluded trials on which an incorrect inference was made.

No participant saw the same cartoon in more than one condition. Thus, even though there were hundred cartoons per condition, each participant received a selection of 25 direct verbal, 25 indirect verbal, 25 direct gestural and 25 indirect gestural. This amounted to 4 stimuli lists. We took care to construct the stimuli lists so that the following variables were balanced for each condition within a given list: the number of words on the first two pictures of each cartoon strip (context), and the number of words and frequency of content words in the reply (for verbal conditions).

2.2 Participants

We recruited 27 participants, but analyzed data from only 24 of them (2 received wrong versions of the experiment and one experienced technical difficulties). They were recruited from the MPI participant database, and were mostly students of the Radboud University in Nijmegen (8 males). All participants provided an informed consent prior to starting the experiment and were paid or awarded course credits for their participation.

2.3 Experimental procedure

Participants were seated behind a computer screen in a sound-proof booth. They read the instructions with example items, and then continued to the cartoons. Each trial began with a 1s fixation cross in the middle of the screen. Next, the two context frames with a context dialog (frame 1) and question (frame 2) were presented simultaneously. The participant was instructed to press a button only after she had fully comprehended what was going on in the pictures. Another 1s fixation cross followed and a target picture with the reply was presented in the middle of the screen. Again, participants were supposed to press the button only after they had understood what the reply *really* meant. The probe sentence followed after a short fixation cross of only 200ms, to prevent them from pressing the button too fast and relying on the pause to finish their inferences. When the probe sentence was presented, the participants decided whether it was a likely/usual continuation or not, and pressed the corresponding button.

The items were presented in 10 blocks, with brief pauses in between them. The entire experiment lasted below an hour.

Each participant also received an IRI questionnaire and a post-test with general questions about their experience, perceived level of difficulty, etc. After finishing the post-tests, the participants were debriefed.

3. Results

The critical event (reply) was defined as time elapsed from target cartoon frame onset until the button-press. We excluded error trials, where participants did not correctly solve the probe task. This happened on 102 trials altogether (4.25%). In some cases, both usual and unusual replies were possible, and we counted both as correct. 3 additional trials were excluded due to other issues (wrong button or skipping the context frame by accident). We also excluded 67 outliers defined as 2.5SD below and above the mean of participant/condition (2.67%). A repeated-measures ANOVA revealed that the outliers were equally distributed among all four conditions (no significant main effects or interactions, all $p>.129$). The same was true for the distribution of errors, although the difference between the number of errors on direct and indirect trials approached significance (41 vs 61 errors, $p=.052$).

We conducted two additional checks on our data. Since the direct and indirect context pictures preceding the target replies were not completely identical in their *verbal* content, we compared their mean reading time. This was to check whether any differences in the speed of target reply interpretation had not arisen already prior to the target reply. There were no statistically significant differences in reading the direct or indirect context pictures (Wilcoxon signed-rank test, $Z=1.429$, $p=.153$). We also conducted a 2x2 (modality and directness) repeated-measures ANOVA on how fast participants answered the post-reply probe sentence. We wanted to exclude the possibility that participants pressed the response key after a superficial reading of the reply, and waited with a deeper interpretation after they had seen the probe sentence. This would have resulted in reaction time differences between conditions after probe questions. It took the participants equally long to answer the probe sentences, regardless of the condition (no main effects or interactions, all $p>.182$). It is therefore unlikely that participants routinely delayed thinking about the meaning of the reply only after they had seen the probe sentence.

The main variable of interest, the latency of interpreting the target reply, was analysed with mixed-effects logistic regression for repeated measures to account for the non-normal distribution of response times, and also for missing values due to errors and outliers. The analysis was carried out with functions from the *nlme* package (Pinheiro, Bates, DebRoy, & Sarkar, 2016) in the statistical environment R (version 3.2.3; R Core Team, 2015), using the Restricted maximum likelihood (REML) estimation approach. Fixed effects were represented by both variables describing the character of the stimuli: Indirectness (with 2 levels – direct/indirect) and Gesturality (verbal/gestural). Random effects were generated by the participants (variable ID) and the particular stimuli (see *Table 1* for full results and *Appendix 1* for model comparisons). Post-hoc analyses were performed using least square means with Tukey's adjustment for multiple comparisons (Lenth, 2016), using R-package *lsmeans*.

Participants interpreted indirect replies slower than direct replies, as evidenced by a main effect of indirectness (Estimate: 84.89, SE: 32.67, $t(2201)=2.60$, $p=.009$). There was no overall modality difference, i.e. gestures were not interpreted slower than verbal utterances in general (Estimate: 82.79, SE: 70.71, $t(2201)=1.17$, $p=.242$). There was, however, a very strong interaction between modality and indirectness (Estimate: 252.31, SE: 59.66, $t(2201)=4.23$, $p<.001$). The difference in reading times between the direct and indirect gestural reply (Estimate: 337.2, SE=49.9, $t(2201)=4.73$, $p<.001$) was almost four times the difference between the direct and indirect verbal reply, which was still marginally significant (Estimate: 84.9, SE: 34.4, $t(2201)=2.60$, $p=.047$) (see *Table 2* for descriptive statistics). Looking at the direct stimuli only, there was no statistically significant difference between the verbal and gestural conditions (Estimate: 82.8, SE: 70.7, $t(2201)=1.17$, $p=.645$). In contrast, there was a large difference between verbal and gestural replies in the two indirect conditions, with gestural indirectness taking on average 335 ms longer than verbal indirectness (Estimate: 335.1, SE: 70.9, $t(2201)=6.76$, $p<.001$). Lastly, a direct comparison between direct gestural and indirect verbal replies showed that they were interpreted at a comparable speed (Estimate: 2.11, SE: 70.79, $t(2201)=0.03$, $p=1$).

Fixed-Effect Parameter	Value	Std.Error	Df	t-value	p-value
(Intercept)	1497.0	90.4	2201	16.6	<.0001
(Indirect)	84.9	32.7	2201	2.6	0.009
(Gestural)	82.8	70.7	2201	1.2	0.242
(Indirect*Gestural)	252.3	59.7	2201	4.2	<.0001

Random Effects	StdDev	Corr
(Intercept)	428.3	
(Gestural)	278.5	0.51
Residual	832.2	

Table 1. GLMM results for the verbal and gestural target replies.

	Direct reply	Indirect reply
Verbal modality	1496.7 (463)	1581.5 (423)
Gestural modality	1579.2 (653)	1917.3 (650)

Table 2. Mean reading times and standard deviations in milliseconds for the target replies in all four conditions.

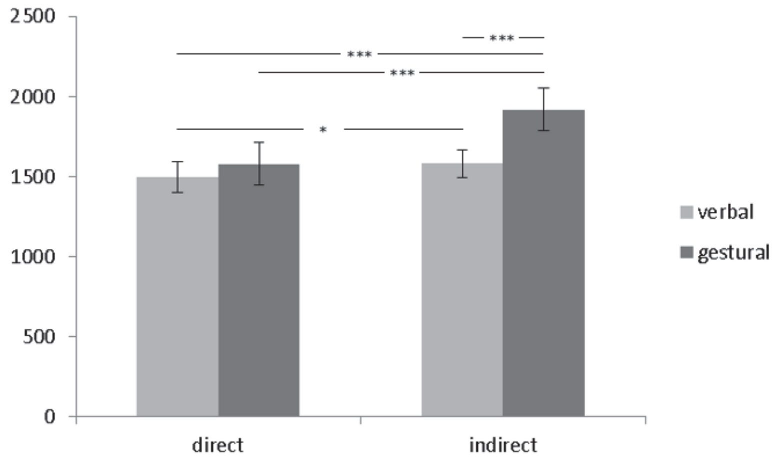


Figure 2: Mean reaction times in milliseconds to target pictures in the four different conditions: Direct and indirect verbal replies and direct and indirect gestural replies. Error bars represent SEM.

In order to see whether individual differences in participants' ability to engage in cognitive empathy influenced their speed of interpreting direct and indirect replies in different modalities, we performed a correlation analysis with speed of target replies' interpretation and scores on the Fantasy (FS) and Perspective Taking (PT) subscales of IRI. Apart from looking at the relationship between individual conditions and these two subscales, we were interested in several other measures. In particular, we computed scores for an overall difference between the two modalities (both gestural – both verbal replies), i.e. how much slower/faster was a participant in the gestural modality in relation to the verbal modality. In the same way, we computed a general score for indirectness (both indirect – both direct replies): how much slower/faster was a participant on indirect replies in comparison to direct replies. Lastly, we looked at modality differences within the two levels of indirectness. That is, how much slower/faster was processing of direct gestural replies relative to direct verbal replies, and indirect gestural relative to indirect verbal replies.

We found that there was indeed a different contribution of the two IRI subscales to measures related to two different aspects of target reply interpretation. Fantasy scale (FS) was related to the interpretation of gestures, rather than indirectness. The higher the participants scored on this scale, the *slower* they were to interpret overall gestural replies relative to verbal replies (r_s (22)=.53, p=.007), as well as direct gestural replies relative to direct verbal replies (r_s (22)=.51, p=.012). There was also a correlation between the FS scale and the speed of interpreting direct gestural replies as such, although only as a trend (r_s (22)=.37, p=.079). These three measures did not correlate with the Perspective taking scale. In contrast, PT was not related to gesture interpretation per se, but rather to indirectness, even though these correlations were weaker and there was only a trend towards significance (p's< .1). The higher participants' PT abilities, the *slower* they were to interpret indirectness in general (r_s (22)=.38, p=.070), as well as indirect gestural replies (r_s (22)=.36, p=.089) and indirect gestural replies relative to direct gestural replies (r_s (22)=.39, p=.062). Again, these measures were not correlated with the FS scale.

4. Discussion

This reaction time study investigated the time-course of communicative meaning comprehension in two different modalities: verbal utterances and pointing/showing gestures, which were either simple (direct) or complex (indirect). The fact that pointing gestures do not contain any encoded information in themselves makes them a perfect test-case for investigating how meaning-making crucially depends on the context. We showed that, even though indirect verbal replies and pointing gestures ultimately lead to the same mental representations and are motivated by the same social intention, indirect gestures seem nevertheless more taxing than their verbal counterparts. The locus of this difference is not a general difficulty interpreting gestures, as there is no corresponding difference between the two direct replies. Rather, as suggested by the pattern of reaction time results and cognitive empathy measures, it is the fact that when context gets less constraining, gestural indirectness becomes more ambiguous.

Based on the literature on pragmatics that differentiates between at least two levels of contextually-supported meaning, we hypothesized that the difference between verbal and gestural indirectness would not be in getting from its explicit to implicit meaning, but rather in identifying the focus of the gesture, or what it is that the communicator is highlighting for her

(the gesture's explicit meaning or "what is said", e.g. Gibbs, 2002). This is not the same as a simple recognition of the objects in the visual scene, because determining the gesture's focus depends on the preceding conversational context and the interlocutors' common ground.
We expected that establishing utterance-level meaning would be comparably easy to compute for direct verbal utterances or direct pointing gestures, since the preceding conversational context was highly constraining. In other words, using language would not have any advantage over gestural signals because both would highlight the relevant aspect of the scene equally efficiently. Language is typically more accurate in describing reality, but linguistic processing is also costly, and in simple contexts, it might be equally cost-effective to use a gesture or a verbal utterance (e.g. pointing to a visually salient threat or exclaiming "Watch out, a car!"). Indeed, we found that there was no difference in the latencies of interpreting the two types of direct speech acts.
In contrast, the two indirect replies should both involve an extra inferential step in order to get to the implicated portions of the message (bridging the gap between explicit and implicit meaning); and, since a gesture is also inherently more ambiguous in highlighting a specific section of reality than verbal utterance, the indirect gestural replies should be more challenging than the indirect verbal replies. This prediction was confirmed, as indirect pointing gestures led to significantly longer response times than indirect verbal replies.

We also predicted that the two interpretative steps (identifying the focus of the gesture and getting from explicit to implicit meaning) would require different kinds of inferential processes. This was partly reflected in the different patterns of correlations with the two cognitive empathy subscales of IRI. We found that processing related to gestures was associated with the Fantasy subscale, while processing related to indirectness, both verbal and gestural, was not related to this subscale. Rather, there was a trend towards association with the Perspective-taking subscale. Intriguingly, we found that participants scoring high on both of these scales were *slower* to interpret gestures and indirectness, respectively. A tentative explanation is that these participants engaged in less superficial processing, e.g. entertained more hypotheses about the gesture's meaning than participants who are low on fantasy and perspective taking, and this was a time-consuming process.
However, these correlational data need to be interpreted with caution because some of these correlations were only marginally significant. Also, the pattern of correlations was not perfect: if FS was related only to identifying the gesture's focus, we would also expect it to correlate

with response times on the indirect gestural items, and if PT was only related to ToM-like pragmatic inferencing, we would expect it to correlate with response-times on the indirect verbal items.

Incidentally, our results do not support another possible explanation, that in order to interpret pointing gestures, one always first needs to verbalize them. Such “verbalization hypothesis” would lead to longer latencies for gestures across the board. However, longest latencies for gestural indirectness are still in line with the possibility that recipients resort to verbalization only when things get more complex, i.e. in case of gestural indirectness. Such account is impossible to rule out with our present data. In the following, we will address some of the remaining issues and limitations of the study.

4.1 Explicit vs implicit judgments of gestures' indirectness

The indirectness pre-test showed that participants judged the direct gestural replies as *less* direct and the indirect gestural replies as *less* indirect than their verbal counterparts. Interestingly, the response times did not replicate this pattern. It seems that this disparity might have been caused by the difference between having to explicitly consider what is direct and what indirect, as opposed to implicit judgments in the actual experiment. We presume that in the indirectness pre-test, participants were taking a different approach to what they considered “literal meaning” or “what is said” in the verbal replies versus in gestures. With the verbal replies, they read the utterance and then took a minimal (direct) or a large (indirect) “mental step” to figure out the actual communicated meaning. Participants' explicit understanding of the literal meaning of the utterance differed between verbal and gestural utterances, and for gestures, it was probably somewhere between “what is said” and “what is implicated”. We base this hypothesis on the replies from the second pre-test where (a different group of) participants had to supply a caption to the gestural items. In most cases, they considered the gesture in the direct gestural replies to already express an *interpretation*. For example, in a situation when somebody could not ride a bike because his hand was in a cast, the verbal interpretation of the gesture supplied by the participants was “I cannot ride with a broken hand”, whereas the indirect verbal caption of the actual item said only “My hand is in a cast”.. Consequently, when asked to explicitly judge how indirect these gestures were, pre-test participants judged them as less indirect than their verbal counterparts because they felt that they were making a smaller “mental step” to get at the communicated meaning. In contrast, the reaction time patterns from the actual experiment which lacked explicit judgments showed that interpreting gestural indirectness was indeed much more time-consuming than verbal indirectness. These results are in line with several studies in the experimental pragmatics literature which show that people generally find it hard to explicitly pinpoint utterance-level meaning (e.g. Nicolle & Clark, 1999).

4.2 Are gestures special?

One question that we cannot answer based on the current results is whether the pointing gesture was necessary in order for the participants to make inferences, or whether just a picture of the relevant object or the situation would allow them to be just as fast, or even faster (see Verbuk & Shultz, 2010). This is complicated by the fact that a carefully constructed cartoon is already a “pointing gesture” of sorts, as the experimenter highlighted only those aspects of the situation she considers relevant. In real life, situations in which pointing occurs are almost always noisier and thus more demanding on part of the recipient. However, our hypothesis is that in real-life situations, a pointing gesture towards an object or a situation contains more information than just the object or situation alone. There are additional clues, also present in the gestural cartoons, most notably the protagonists’ facial expressions accompanying the gestures. This could have helped the participants to not only locate the relevant object/event (which would also be possible based on just presenting a picture of the relevant part of reality, e.g. a worn-out running shoe) but presumably also sped up their inferences by understanding what the protagonist’s social intention is, i.e. her attitude towards the object/situation. For instance, in the running shoe item, the protagonist’s face indicated that he feels bad about the situation, which might have helped the participants to be faster in inferring that his friend is declining his invitation (“*Will you join me for a run?”*). Again, exactly how such facial expressions can help in interpreting communicative gestures is a question for further research.

It is also clear that having someone point to objects or situations in real life leads to substantially different mental operations than noticing these objects on one’s own. The fact that someone cooperatively picks out and highlights a section of reality for their conversational partner means that the recipient needs to consider speaker-related information: what is in their shared common ground, and what are the speaker's conversational and wider social goals. In other words, a pointing gesture towards an object as an ostensive signal vehicles the communicator's communicative intention, whereas simply seeing an object or an image of an object only carries informative intention (see also Csibra, 2010). One piece of evidence suggesting that there indeed might be a substantial processing difference comes from a study by Peeters, Snijders, Hagoort, & Özyürek (2017), who showed that there was a difference in brain activation when a referentially salient object was highlighted via a pointing gesture or just a visual cue.

4.3 Pointing gestures in real-life situations

With respect to exploring gestural indirectness in more detail, there are several other factors worth considering. For one, when people use gestures in the real world, it is often situationally motivated. For example, they use pointing to accompany a demonstrative (this/that), or their mouth is full, the room is noisy, they do not have sufficient command of the target language, or they want to convey something "behind someone's back", like in our initial example with children and taboo words. On the part of the recipient, such situational cues might help guide interpretation, since they signal *motives* for why the speaker uses gestures instead of speech. It was not possible to enact such situations in simple, static cartoons, but further research might use video clips to achieve this and explore whether interpreting the gesture's meaning will get easier relative to cartoons that do not give any clear motivation for why characters communicate gesturally.
Using more dynamic stimulus material might be also useful for exploring another aspect of communication – facial expressions acting as pointers to the speaker's social intention, presumably in a similar way as intonation and prosody in verbal utterances. We did try to account for the importance of facial expressions by drawing them as accurately as possible, but they were still only schematic in comparison to real-life situations.

Further research might also want to explore composite verbal-gestural signals. There has been some research on the role of gesture in pragmatic processing. Kelly (2001) and Kelly et al. (1999) showed that gestures help with processing of pragmatic meaning, both in adult and children.

5. Conclusion

Recipients are generally good at interpreting utterances whose conveyed meaning is very different to their literal or explicit meaning, and clearly, this process is impossible without taking the various levels of context into account. Our paradigm manipulated context-depended meanings in two ways: apart from needing to work out the implicit portion of the utterances, listeners also needed to work out their explicit meaning, especially when processing pointing or showing gestures. Such indirectly used gestural utterances carry no meaning in themselves but fully derive it from the visual scene, mutually known common ground, tacit rules governing dialogic exchanges and immediate conversational context set up by the preceding exchange. Our results, based on the pattern of reading times and correlations with different dimensions of cognitive empathy, suggest that there might be at least two classes of inferential processes behind gestural indirectness. An fMRI study using the same stimuli and design might be able to provide further confirmation and refinement of this idea, and provide details about the specific inferences involved.

Direct context story	Direct target reply/gesture	Indirect context story	Indirect target reply/gesture
Two girls are on a night out, standing in front a pub. A says that this is her ex's favorite pub and that she really does not want to encounter him tonight. B says that they should rather go someplace else because A's ex is here. A asks how does she know.	His bike is here. Points to a male bike parked outside of the pub.	Two girls are on a night out, standing in front a pub. A says that this is her ex's favorite pub and that she really does not want to encounter him tonight. B asks whether she thinks he is inside.	His bike is here. Points to a male bike parked outside of the pub.
A wants to get to the airport and asks B for a ride. B says he cannot help him out. A asks why is that.	I've had some wine. Points to a half-empty wine glass nearby.	A wants to get to the airport and asks whether B can give him a ride in his car.	I've had some wine. Points to a half-empty wine glass nearby.
Two friends in a restaurant. A asks whether they should order an ice-cream for dessert. B declines. A wants to know the reason.	I have a toothache. Points to her jaw with a pained expression on her face.	Two friends in a restaurant have just finished dinner. A asks whether they should order ice-cream for desserts.	I have a toothache. Points to her jaw with a pained expression on her face.
A and B talking about their pet cat who is ill. A asks whether he feels better and B replies that he still has no appetite. B asks how does A know that.	His bowl is still full Points to a bowl full of cat food.	A and B talking about their pet cat who is ill. A asks whether he feels better and whether he has finally eaten.	His bowl is still full Points to a bowl full of cat food.
Two old friends meet at a party after a long time. A asks whether he can fetch B a drink, and she specifies that she wants something non-alcoholic. A is curious why.	I am pregnant. Points to her pregnant belly (which has been hidden behind a purse on the previous picture.)	Two old friends meet at a party after a long time. A asks whether he can fetch B a drink of wine.	I am pregnant. Points to her pregnant belly (which has been hidden behind a purse on the previous picture.)
A wife and a husband are watching TV. The husband asks whether he can turn up	The baby is asleep.	A wife and a husband are watching TV. The husband asks whether he can turn	The baby is asleep.

the volume but the wife says no. He wants to know the reason.	Points to a baby sleeping in a nearby cot.	up the volume.	Points to a baby sleeping in a nearby cot.
Two friends are getting ready to go out, A looks out of the window and says that it is raining. B asks how can they protect themselves from the rain.	I have an umbrella. Shows up an umbrella and points to it.	Two friends are getting ready to go out, A looks out of the window and says that it is raining. B asks whether they should stay at home instead.	I have an umbrella. Shows up an umbrella and points to it.
A comes into the office and asks his colleague, B, whether she wants to join him for lunch. B says that she cannot go and A wants to know the reason.	I have lots of work to do. Points to a huge pile of documents on her desk.	A comes into the office and asks his colleague, B, whether she wants to join him for lunch.	I have lots of work to do. Points to a huge pile of documents on her desk.
Two friends, dressed in fancy clothes, are going home after a night out. They are at a bus stop and have just missed the bus. A asks B whether they should just walk home. B says no and A wants to know why.	I have high heels on. Points to her high-heeled shoes.	Two friends, dressed in fancy clothes, are going home after a night out. They are at a bus stop and have just missed the bus. A asks B whether they should just walk home.	I have high heels on. Points to her high-heeled shoes.
An older couple is sitting at a park bench. The man wants to light up a cigarette but the lady asks him not to. He wants to know why.	Kids are playing nearby. Points at kids playing nearby.	An older couple is sitting and relaxing at a park bench. The man asks the lady whether he can light up a cigarette.	Kids are playing nearby. Points at kids playing nearby.

Table 3. Examples of stimuli used in the study. Context stories are descriptions, not actual texts of the dialogs. The structure was always the same: opening dialog where both A and B had a line on frame 1; question from one of the characters on frame 2; reply by the other character via a sentence or a gesture on frame 3.

References

Bašnáková, J., van Berkum, J., Weber, K., & Hagoort, P. (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers? *Neuropsychologia*, *76*, 79–91. https://doi.org/10.1016/j.neuropsychologia.2015.03.030

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the language given: the neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578. https://doi.org/10.1093/cercor/bht112

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brown, P., & Levinson, S. C. (1987). *Politeness: some universals in language usage*. Cambridge University Press.

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., … Aureli, T. (2014). Do You Know What I Mean? Brain Oscillations and the Understanding of Communicative Intentions. *Frontiers in Human Neuroscience*, *8*, 36. https://doi.org/10.3389/fnhum.2014.00036

Bucciarelli, M., Colle, L., & Bara, B. G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, *35*(2), 207–241. https://doi.org/10.1016/S0378-2166(02)00099-1

Carston, R. (2013). 12. Implicature, explicature and truth-conditional semantics. *The Semantics-Pragmatics Boundary in Philosophy*.

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2011). A word in the hand: action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1585).

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*(3), 567–598. https://doi.org/10.1080/01638539709545003

Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language*, *25*(2), 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

Cutica, I. (1980). Neuropsychological Evidence for Linguistic and Extralinguistic Paths in Communication Comprehension: Understanding the Actor's Communicative. *Brain*, *27*(27), 518–523.

Cutica, I. (2005). Neuropsychological Evidence for Linguistic and Extralinguistic Paths in Communication. In *Proceedings of the Cognitive Science Society* (Vol. 27).

Cutica, I., Bucciarelli, M., & Bara, B. G. (2006). Neuropragmatics: Extralinguistic pragmatic ability is better preserved in left-hemisphere-damaged patients than in right-hemisphere-damaged patients. *Brain and Language*, *98*(1), 12–25. https://doi.org/10.1016/j.bandl.2006.01.001

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2016). Brain basis of communicative actions in language. *NeuroImage*, *125*, 857–867. https://doi.org/10.1016/j.neuroimage.2015.10.055

Enfield, N. (2009). The anatomy of meaning: Speech, gesture, and composite utterances.

Enfield, N. J., Kita, S., & de Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722–1741. https://doi.org/10.1016/j.pragma.2007.03.001

Garrett, M., & Harnish, R. M. (2007). Experimental pragmatics: Testing for implicitures. *Pragmatics & Cognition*, *15*(1), 65–90. https://doi.org/10.1075/pc.15.1.07gar

Gibbs Jr., R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, *34*(4), 457–486. https://doi.org/10.1016/S0378-2166(01)00046-7

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax And Semantics* (Vol. 3, pp. 41–58). Academic Press.

Grice, H. P. (2008). Further notes on logic and conversation. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of Human Inference and Its Foundations*. Cambridge University Press.

Holtgraves, T. (1998). Interpreting indirect replies. *Cognitive Psychology*, *37*(1), 1–27.

Karttunen, L., & Peters, S. (1975). Conventional Implieature in Montague Grammar. *Proc. of the First Annual Meeting of the Berkeley*.

Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language*, *28*(2), 325–349. https://doi.org/10.1017/S0305000901004664

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a Hand to Pragmatic Understanding: The Role of Speech and Gesture in Comprehension and Memory. *Journal of Memory and Language*, *40*(4), 577–592. https://doi.org/10.1006/jmla.1999.2634

Kendon, A. (2004). Gesture: Visible action as utterance.

Kissine, M. (2015). Pragmatics as Metacognitive Control. *Frontiers in Psychology*, *6*, 2057. https://doi.org/10.3389/fpsyg.2015.02057

Kita, S. (Ed.). (2003). *Pointing: Where Language, Culture, and Cognition Meet*. Psychology Press.

Lenth, R. (2016). Least-squares means: the R package lsmeans. *J Stat Softw*.

Nicolle, S., & Clark, B. (1999). Experimental pragmatics and what is said: a response to Gibbs and Moise. *Cognition*, *69*(3), 337–54.

Peeters, D., Snijders, T. M., Hagoort, P., & Özyürek, A. (2017). Linking language to the visual world: Neural correlates of comprehending verbal reference to objects through pointing and visual cues. *Neuropsychologia*, *95*, 21–29. https://doi.org/10.1016/j.neuropsychologia.2016.12.004

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2016). *R Core Team. nlme: linear and nonlinear mixed effects models. 2014. R package*.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the*

National Academy of Sciences of the United States of America, *105*(3), 833–838. https://doi.org//10.1073/pnas.0707192105

Quine, W. (1960). Word and object (Studies in Communication). *New York and London: Tech-Nology Press of MIT*.

Recanati, F. (2004). *Literal meaning*. Cambridge University Press.

Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3-Year-Old Children Make Relevance Inferences in Indirect Verbal Communication. *Child Development*, *84*(6), 2079–2093. https://doi.org/10.1111/cdev.12093

Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition*, *136*, 91–98. https://doi.org/10.1016/j.cognition.2014.11.036

Sperber, D., & Wilson, D. (1995). Relevance: communication and cognition. *Behavioral and Brain Sciences*, *2nd*(4), 697–710. https://doi.org/10.1017/S0140525X00055345

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Verbuk, A., & Shultz, T. (2010). Acquisition of Relevance implicatures: A case against a Rationality-based account of conversational implicatures. *Journal of Pragmatics*, *42*(8), 2297–2313. https://doi.org/10.1016/j.pragma.2010.01.005

Walker, T., Drew, P., & Local, J. (2011). Responding indirectly. *Journal of Pragmatics*, *43*(9), 2434–2451. https://doi.org/10.1016/j.pragma.2011.02.012

Wilson, D., & Sperber, D. (1986). Pragmatics and modularity. *In: Farley, A and Farley, T and McCullough, K, (Eds.) (Proceedings) Chicago Linguistic Society 22. (Pp. Pp. 67-84). Chicago Linguistic Society: Chicago, IL. (1986)* .

Wilson, D., & Sperber, D. (2004). Relevance Theory. In G. Horn, L.R. & Ward (Ed.), *The Handbook of Pragmatics* (pp. 607–632). Oxford: Blackwell.

CHAPTER 5: NEURAL CORRELATES OF GESTURAL INDIRECTNESS

I see your point: the neural correlates of understanding indirect communicative pointing gestures

Jana Bašnáková[1,2], Jos J.A. van Berkum[3], Emanuela Campisi[4], Peter Hagoort[1]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands [2] Institute of Experimental Psychology, CSPS SAS, Slovakia [3] UiL-OTS, Department of Languages, Literature and Communication, Utrecht University, Netherlands [4] Department of Humanities, University of Catania, Italy

Abstract

Communication is not limited to the linguistic code: conversations often rely on signals that are both non-conventional and non-verbal, such as pointing gestures. In an fMRI study, we compared the neural correlates of speaker meaning comprehension from verbal utterances and matched pointing/showing gestures. To vary the inferential load, half of the utterances were direct and half indirect. In this way, we could investigate which brain regions are implicated in processing utterances that lack any coded meaning (pointing gestures), and whose interpretation involves bridging the gap between what the speaker says, and her intended message (indirect speech). Our hypothesis was that indirectness should engage the same mentalizing regions regardless of modality, since it expresses the same social intentions. We found that the difference between verbal and gestural indirectness was indeed not detected in regions associated with social cognition (medial prefrontal cortex, temporo-parietal junction, precuneus), which responded equally strongly to all indirect replies. Instead, the differential effect was in the right BA45 in the inferior frontal gyrus which is also seen in studies on interpreting non-conventional signs. We conclude that while the intentional–inferential infrastructure for communication is the same regardless of the signal's modality, the ambiguity inherent in indirect pointing gestures makes it more difficult to establish their utterance meaning, relative to verbal indirectness.

Based on:

Jana Bašnáková, Jos van Berkum, Emanuela Campisi & Peter Hagoort (in preparation). I see your point: the neural correlates of understanding indirect communicative pointing gestures.

1. Introduction

Communication can be a lot more than conventional and coded, as almost anything will, under the right circumstances, become a communicative signal: an outstretched index finger, hint of a smile, or even a well-timed silence. Examples like these make it evident that the cognitive architecture enabling human communication is not limited to operations with stored word meanings and syntactic rules. The crucial ingredient is the ability to recognize the speaker's intention, or what the speaker wants us to know, think or do, when she ostensively performs a specific communicative signal (Grice, 1975; Sperber & Wilson, 1995; Tomasello, 2008; Wilson & Sperber, 1986).

There are at least three levels of intentions at play, as Tomasello (2008) demonstrates with an example of pointing gestures. When a friend points to a visual scene, such as a cat standing next to her empty bowl, we must first recognize that she has a *communicative intention*: that she is stretching her index finger out *for us*, to get a certain message across, and not just for herself, to examine the cracks in her nail polish. As the recipients, we must also recognize our friend's *referential intention*, or what specifically is she drawing our attention to. This is, however, still not enough to understand her message: getting what she means is closely related to recognizing her *social* intention, or *why* is she drawing our attention to that scene. She might be pointing because she wants us to notice the cat's elegant posture, because she wants us to realize that the bowl is empty, or because she wants us to *do* something about the emptiness of the bowl.

How do recipients figure out the speakers' intentions? Once the recipient recognizes that the speaker performs a communicative signal *for him*, it raises specific expectations of the signal's relevance, motivating inferences that ultimately lead to the interpretation of the speaker's message (Grice, 1989; Sperber & Wilson, 1995; Wilson & Sperber, 1986). These inferences crucially depend on the context in which the communication is taking place, including information that is *mutually shared* between the speaker and the listener, such as their world-knowledge, personal history, previous discourse, or the current setting (Clark, 1996; Clark & Carlson, 1982; Clark & Marshall, 1981). Depending on what is in this common ground, even physically identical communicative signals can give rise to very different social intentions. By pointing to the cat, the speaker can simply inform the recipient about her feeding habits: *"Look, the cat has eaten everything"*. Alternatively, if the recipient usually feeds the cat at around this time, but forgot about it today, then the speaker's pointing gesture to the cat's empty bowl might be a request: "*Please, feed the cat*". In the first case, the speaker's referential intention (*the cat's bowl contains no food*) and her social intention (*I am informing you that the cat has eaten it up*) have a fairly straightforward relationship. In the second case, however, the relationship between the speaker's referential intention (*the cat's bowl contains no food*) and her social intention (*I am requesting that you to fill it up*) is much less direct. The recipient must, therefore, bridge a greater inferential distance between what the speaker indicated by her gesture, and what message she actually intended to convey.

In this study, we explore the neural systems that support the ability to make such inferences from signals in two different modalities: pointing gestures or language. By comparing the processing of indirect replies expressed in different modalities, we aim to show to what extent the pragmatic inferences behind speaker meaning interpretation in these two different modalities are supported by the same neural systems, and in what respects they might differ. Before describing our experimental paradigm and predictions, we discuss the relevant prior research on communicative pointing, verbal indirectness, and the combination of pointing and indirectness.

1.1 Understanding communicative pointing

Even though there are now many neuroimaging studies on gesture comprehension (see Yang, Andric, & Mathew, 2015, for a meta-analysis), only a handful of them examine *pointing* gestures used in communicative settings. Brunetti et al. (2014) used MEG to compare producing and interpreting communicative pointing to request things (imperative pointing) with pointing to share attention to an outside entity (declarative pointing). Declarative pointing involves sharing attention between the listener and the communicator, and therefore requires that the listener attends to the communicator's mental state, whereas imperative pointing only has an instrumental goal that can be inferred directly from the agent's overt behavior (Brunetti at al., 2014). Both comprehension and production of declarative pointing, relative to imperative pointing, modulated activation in regions associated with social cognition: the medial frontal cortex (dorsal anterior cingulate cortex (ACC)), and the right posterior superior temporal sulcus (STS). In an adaptation of the same task for fMRI (Committeri et al., 2015), the production and comprehension of declarative pointing elicited higher activation than imperative pointing in the medial frontal cortex (MFC), middle insula, ventral premotor cortex and the right pre-supplementary motor area (SMA).

There are two more studies which looked at pointing gesture comprehension in communicative settings, although their gestural conditions always included other types of gestures alongside pointing, e.g. emblems or pantomime. In an fMRI study by Redcay, Velnoskey, and Rowe (2016), the participants saw an actress performing a series of two communicative gestures (*point forward + tap ear*) directed at them, or say a sentence in the second-person matched to the content of the gestures ("*I want you to listen to me*"). In the control conditions, there were two non-communicative self-adaptor gestures (*smooth hair +pull shirt*), as well as non-matched, standard third-person sentences ("*He wore a sweater to keep warm*"). A conjunction of meaningful and communicative gestures and sentences identified the left inferior frontal gyrus (IFG) and bilateral middle STS, extending posteriorly on the left and anteriorly on the right.

Lastly, Enrici, Adenzato, Cappa, Bara, and Tettamanti (2011) compared communicative intentions expressed via various types of gestures, or via written sentences, presented in cartoons (two people seated at a table with a bottle of wine; verbal condition – *"Please, pass me the bottle"*, gestural condition – *one of them points to the bottle*), with a control condition of physical causality (after two context frames showing a shelf that is about to fall down due to a crack in the wall; verbal – *"The shelf falls down"*, gestural – *a picture of the shelf falling down*). Intention processing in communication was not fundamentally affected by the chosen expressive means – there was a set of core regions, including the precuneus, medial prefrontal cortex (MPFC), bilateral STS and temporo-parietal junction (TPJ), involved in communicative intention interpretation regardless of whether it was delivered via gestures or written sentences.

To sum up, these neuroimaging studies showed that interpreting communicative, direct pointing or other gestures frequently activated regions that belong to what Bara, Enrici, & Adenzato (2015) call the *intention processing network* (IPN), consisting of the medial prefrontal cortex, bilateral precuneus and TPJ /posterior STS. Bara and colleagues hypothesize that intentions in general engage some of the nodes of this network, but that only *communicative* intentions activate the entire network.

1.2 Understanding indirect language

Everyday utterances are hardly ever entirely explicit, and there is almost always a need to consider some contextual information apart from what is “in” the signal (e.g. Bach, 1994; Carston, 2008; Garrett & Harnish, 2007; Kissine, 2015). In many instances, what the speaker actually says - or, the level of meaning standardly referred to as “what is said” in pragmatics (e.g. Garrett & Harnish, 2007), does not correspond to, or is even directly opposite, to what the speaker really means to convey (or “what is meant/implicated”) (e.g. Recanati, 2004). For the purpose of this paper, we can approximately map these distinctions onto Tomasello’s levels of intentions. “What is said” would correspond to the speaker’s referential intention (*what* is the speaker directing my attention to), and “what is meant” to the speaker’s social intention (*why* is she directing my attention to this, what does she want me to know/feel/do).

Indirect utterances are a good example of a large distance between what the speaker says and what she really means. People use indirect means of expression for diverse reasons. For example, as a face-saving strategy to maintain politeness in interactions (Brown & Levinson, 1987; Holtgraves, 1998), with interlocutors in many cultures avoiding expressing criticism directly (“*It’s hard to give a good presentation*,” instead of “*Your presentation was terrible.*”). Another reason for being indirect is to achieve plausible deniability (Pinker, Nowak, & Lee, 2008), i.e. to say things in a manner that the speaker’s ulterior motive is signalled to the listener, but that she can still avoid being held directly accountable (a boss telling her subordinate “I *hope* you can do” a potentially illegal act instead of *directing* him to carry out such act).

Yet another, less emotionally loaded, grounds for using indirectness is “information economy”, the efficient packaging of information into a single utterance (Walker, Drew, & Local, 2011). As for the latter, sometimes speakers reply to yes/no questions in a way that helpfully pre-empts further questioning, addressing what they thought was the *motivation* behind the question in the first place. For example, when a student asks his housemate whether he wants to join him for a run, the housemate can choose a direct reply, “*No, I will not*”. Such a reply would, however, be less informative than answering indirectly by detailing the reasons why he cannot go (“*My running shoe is worn out*”). In situations like these, indirect answers require the listener to do some inferencing in order to get from the speaker’s referential intention (*His shoe is worn out*) to his social intention (*He is declining my offer* *because his shoe is worn out*). At the same time, indirectness will save the listener the effort necessary to follow up the original question to get at the information that he needs.

Indirect speech acts have been explored in several neuroimaging studies in the recent years (Van Ackeren, Casasanto, & Bekkering, 2012; Bašnáková, Van Berkum, Weber, & Hagoort, 2015; Bašnáková, Weber, Petersson, Van Berkum, & Hagoort, 2014; Feng et al., 2017; Jang et al., 2013; Shibata, Abe, Itoh, Shimada, & Umeda, 2011), including our own work on mostly face-saving indirectness. The emerging picture is that working out the social intention behind the utterance engages both the bilateral fronto-temporal language areas associated with processing language in context, such as inferior frontal gyrus and middle/superior temporal gyrus; as well as regions associated with mentalizing or intention processing, primarily the temporo-parietal junction and medial prefrontal cortex.

1.3 Understanding indirectness expressed via pointing and language

There are at least two explicit claims in the literature which predict that the intentional-inferential infrastructure behind verbal and non-verbal indirect meaning recovery should be identical, regardless of whether the starting point was given by verbal code (e.g., saying *"the bowl is empty"*) or a pointing gesture (e.g., *pointing to an empty bowl*). The cognitive pragmatics account of Bara and colleagues (Bara et al., 2015; Bucciarelli et al., 2003; Bucciarelli et al., 2006) predicts that if a verbal and a gestural expression lead to the same representations in the mind of the recipient, and if the inferential load to get from their literal to their communicated meaning is the same, modality should not play a decisive role. The same conclusion is reached by Tomasello (Schulze & Tomasello, 2015; Tomasello, 2008; Tomasello, Carpenter, & Liszkowski, 2007), whose approach is built on developmental considerations, and presumes that being able to produce and comprehend pointing gestures by children requires essentially the same socio-cognitive skills as, later in life, producing and comprehending spoken language. Relative to direct pointing, interpreting *indirect* pointing involves making a larger inferential step from the speaker's referential intention to her social intention, and comprehending indirect speech involves, presumably, an equally large step, even though the referential intention is expressed in the linguistic code.

To date, there are only a handful of studies directly testing these hypotheses with indirect speech acts. A pediatric behavioral study by Bucciarelli, Colle, & Bara (2003) compared indirect communication in the verbal and gestural modalities, and their results supported claims that interpreting verbal and gestural indirectness is equally difficult. The authors assessed the comprehension of speech acts of varying complexity in children between 2.6 to 7 years of age, depicted in short videos. Among these, they included direct and indirect utterances, which were either conventional (*Person A draws person's B attention and points to a window she wants to close* / A: *"Sorry, could you close the window?"*) or non-conventional (*Person A points to a window shop for an attractive toy, person B shows up her empty purse / A:"Would you get me that toy?", B:"We don't have any money."*). As expected, non-conventional indirect speech acts were more challenging than conventional ones. Crucially, though, there were no modality differences in the number of errors children made on verbal or gestural indirect speech acts, suggesting that the mode of delivery did not play a decisive role in how well they were able to understand indirectness. A later study on adult right-hemisphere-damaged (RHD) patients and healthy controls using a subset of these speech acts, however, replicated these result only partly. Cutica (2005) found that the performance of RHD patients was worse on gestural than on verbal speech acts across the board, including (non-conventional) indirect and direct ones. Control participants, who are more representative of the general population, also showed some modality differences, although not on all speech acts: they performed worse on indirect speech acts expressed gesturally than on their verbal equivalents, and equally well on the verbal and gestural versions of direct speech acts and ironies.

Our own findings from a reaction-time study with adult participants (Bašnáková, Van Berkum, Campisi, Mentel, & Hagoort, submitted) suggest that while there is very likely substantial equivalence in the intentional-inferential infrastructure necessary to interpret matched indirect speech acts delivered via verbal code or a pointing gesture, there also seem to be interesting differences. We used cartoons which ended with a verbal utterance or a pointing gesture that were – relative to the preceding question - either direct or indirect. For simple, direct replies (A: *"Why can't you join me for a run?",* B:*"My shoe is worn out."/Points to a worn-out shoe*), it indeed did not matter which modality they were expressed in – both verbal and gestural replies took equally long for the readers to grasp. Also interpreting the meaning of indirect communication (A:*"Will you join me for a run?",* B:*"My shoe is worn out."/Points to a worn-out shoe*) took, as expected, more time than interpreting direct communication. We attributed this delay to bridging a larger gap between the speaker's referential and social intentions, which required, for example, taking into account more complex contextual information. The key finding was a highly significant difference between indirectness in the two modalities, with gesturally expressed indirect replies taking the longest time of all: the delay caused by indirectness in pointing was almost four times as big as that between direct and indirect verbal replies.

In our interpretation of this delay, we referred to differences at the level of the referential intention of the speaker – i.e. what does the gesture single out in the environment. We presumed that, when faced with a pointing gesture, listeners struggle with establishing what exactly is the speaker directing their attention to (the gestural equivalent of "what is said"). A pointing gesture, or the piece of reality it singles out, is fairly ambiguous, and thus essentially becomes a non-conventional sign, the meaning of which has to be worked out. This inferential process becomes more demanding when there is a large possible space of choices – such as when the question requires a yes/no answer, but the "yes" or "no" has to be inferred from the reply itself. Therefore, the recipient has to determine not only *how* is that aspect of reality relevant to the yes/no answer, but also *which part* of it is relevant (*Is it the existence of the shoe? Is it its color, or its brand? Is it the fact that the shoe is not in a good shape? Or is the speaker showing me that he just bought these shoes? Or that he is busy because he needs to buy new shoes?*). The indirect verbal reply is different in this respect, because even in less constrained contexts, language packs more "into the signal" than a pointing gesture (Tomasello, 2008), and is therefore more helpful at singling out specific aspects of situations (e.g. Tylén, Weed, Wallentin, Roepstorff, & Frith, 2010).

Thus, while in both indirect situations, the recipient must figure out what the ostensive communicative signal implies (i.e. the speaker's social intention), there is an added level of uncertainty in the indirect gestural case because it is unclear what exactly is being picked out for him. The combination of these two sources of uncertainty was presumably reflected in the long interpretation time for our indirect gestural replies.

In this fMRI study, we explore this issue further, using the same design. The added value of knowing what brain regions are associated with verbal and gestural indirectness is that we can be more specific about the identity of the cognitive processes involved. Our findings will be informative in two ways. First of all, to the extent that the pragmatic inferences to get from the referential to the social intention behind indirectness are indeed identical across the verbal and gestural modality, we expect that both verbal and gestural indirect replies will engage areas involved in this process to the same extent. Based on our previous studies on verbal indirectness, (Bašnáková et al., 2014; Bašnáková et al., 2015), we expect that, relative to interpreting direct replies, interpreting indirectness activates mentalizing areas (MPFC, TPJ, precuneus, see also Bara et al., 2016) to a larger extent, because the listener has to invest more effort in considering the speaker's thoughts, beliefs and motivation. In addition, because a broader range of contextual information needs to be drawn upon in order to cross the referential-social intention distance, we expect that indirectness will also engage discourse-related areas, such as MTG, and bilateral inferior frontal regions, more than direct replies. Secondly, the results will be informative as to whether there is, as predicted by our behavioural study, indeed a differential brain response between indirect verbal and indirect gestural replies in region(s) involved in working out the recipient's referential intention.

Regarding the identity of these regions, our interpretation of the behavioural results suggested that part of the problem with establishing what the pointing gesture singles out is that the signal itself does not give us immediate cues about how the indicated visual scene is relevant to the interaction, or which part of it is relevant. In essence, we can think about pointing gestures as non-conventional signs (i.e. a particular visual scene singled out by the gesture), the meaning of which must be established by drawing on contextual information. A study by Tylén, Wallentin, & Roepstorff (2009) suggested that the right pars triangularis (BA45), part of the inferior frontal gyrus, is a region primarily involved in establishing meaning of such nonconventional communicative devices. Tylén et al. focused on the communicative significance of material objects, or static scenes made of material things arranged in a way that they would signal social meaning, e.g. chairs put out on the street to reserve a parking place vs chairs conventionally placed around a table. In this nonverbal, passive-viewing task, they found that the right pars triangularis was modified by the perceived degree of conventionality of scenes with an intrinsic communicative meaning (chairs as place-holders). The authors concluded that the involvement of the right IFG reflects the pragmatic and contextual integration necessary to resolve ambiguous or nonconventional signs; in other words, when the meaning of the material signals was not intrinsic to the scene, it required additional contextual cues in order to decide among multiple semantic information, or "senses" in which the sign could be relevant (Tylén et al., 2009).We propose that this is exactly the issue that will arise for gestural, but not for verbal, indirectness. The visual scene highlighted by a pointing gesture can have multiple meanings which will need to be

determined on the basis of contextual information.
However, with regards to terminology, we do not agree with Tylén et al. that this problem is "semantic" in nature: this would involve several conceptual representations related to a single word-form. This cannot be the case with gestures, as there is no word form attached to the pointing gesture or the entire visual scene. Rather, we choose to refer to this as "ambiguity in referential intention" of the speaker (or "referential ambiguity" as shorthand): when faced with a visual scene highlighted by the pointing gesture, the addressee must figure out which of the multitude of senses associated with the shoe is currently the relevant one (is it the entire shoe, is it its color, or the hole? Is it some other aspect of the shoe, e.g. that it is new and therefore cannot be run in?).

Is there a similar puzzle to be solved also in case of direct pointing gestures? The ambiguity in the referential intention is not likely to be an all-or-nothing phenomenon. There might be instances where a direct pointing gesture picks out the referent perfectly; for example, when A asks which one of the two running shoes is damaged, then a pointing gesture from B is just as good in indicating the intended referent as saying "*The left one*". In contrast, when A asks "*What is wrong with your shoe?*", and B points to a sneaker with lots of wear and tear, B might still be unsure as to which particular damage A refers to. However, no matter how difficult it will be to zoom in on the exact referent in the direct case, the contextual expectations arising from the relatively clear speaker's social intention (and the preceding context, e.g. a request for a particular piece of information) should always be stronger and thus place more constraints on the pool of possible referents. Therefore, we predict that the "ambiguity problem" might arise for direct pointing gestures as well, but should be considerably smaller than for indirect gestures.

1.4. Experimental paradigm and predictions

We used a series of cartoons displaying interactions between various pairs of protagonists, concluded with an indirect or a direct target reply delivered via a pointing/showing gesture or a written sentence by one of the characters (for further information on the paradigm, see also Bašnáková, van Berkum, Campisi, Mentel, & Hagoort, submitted , see also *Table 1* and *Figure 1*). Importantly, the content – or the "utterance meaning" - of the target reply was identical for the two verbal and two nonverbal conditions, and as closely matched between modalities as possible.

Figure 1. Example of a stimulus used in the experiment: A) direct conditions, B) indirect conditions. The first two frames constitute context and were presented together. The last, target frame, was presented on its own. After the participant indicated that she understood by pressing a button, a true/false probe question appeared on the screen (e.g. *The boys are now going for a run together.*) The stimuli were presented in Dutch.

The participants were asked to read through the cartoons and press a button as soon as they understood the reply. A simple probe question about a possible continuation of the situation was presented after each cartoon (*The boys are now going for a run together/Only one of the boys will now go out for a run*), to keep the participants focused, and to be able to tell whether they interpreted the replies in the way we had intended. For more item examples, see *Table S2*.

	Verbal	Gestural
Direct (*Why can't you go for a run?*)	*My shoe is worn out.*	Holds up a worn out shoe and points to a hole in the sole.
Indirect (*Would you like to join me for a run?*)	*My shoe is worn out.*	Holds up a worn out shoe and points to a hole in the sole.

Table 1. Examples of the four experimental conditions

Summing up, we have the following predictions:

1.) Brain regions involved in interpreting the social intention behind indirectness will be identical for verbal and gestural signals. We expect to find a main effect of indirectness (both indirect > direct replies) which should include the core mentalizing areas (MPFC, bilateral TPJ) and precuneus, as well as regions involved in discourse processing/contextual integration, most notably, bilateral IFG (BA47) and right temporal areas; crucially, we do not expect any interaction with modality within mentalizing areas.

2.) Regarding our second prediction, we expect to find an interaction between indirectness and modality in the right BA45, which has been suggested to be involved in interpreting nonconventional signals (Tylén et al., 2009). This region has been identified on the basis of the results of our previous behavioral study. Here, we predict significantly higher activation for gestural than for verbal indirectness. For this ROI, we also selected two sets of control regions: its left homologue, as well as the bilateral BA47, a neighboring region within the IFG. The bilateral BA47 was chosen because we expect it to be involved in indirectness, but we do not expect modality-specific involvement. This region is known to be involved in situation-model building (e.g. Menenti, Petersson, Scheeringa, & Hagoort, 2009), and since it is a relatively late step in meaning processing, it should be equally difficult for verbal and gestural indirectness.

2. Materials and Methods

The materials were identical to the ones used in our behavioral study (Bašnáková et al., in preparation). The experimental items consisted of 400 cartoon strips, with two *context* pictures and one *target* picture (see *Figure 1*). On the first context picture, two characters talked about something (e.g. *going for a run*), and on the second one, one of the characters asked a question (e.g. "*Do you want to join me?"*). Then, on the target picture, the second character replied. While the context pictures were always the same with regard to the drawing, there were two versions in terms of the accompanying dialog: a direct-reply one and an indirect-reply one. In the direct condition, the question was always a "Wh-" or "How" question. For example, "Why don't you want to go for a run?". In the indirect condition, the question was a yes/no question, for example "Do you want to join me?". The overall length of the text in the context pictures (number of words and syllables), as well as semantic similarity of the context text to the verbal target reply were balanced for the direct and indirect conditions.

The reply, given on the target picture, had four different versions, corresponding to the four experimental conditions. The two verbal conditions (direct and indirect) showed the character's face and part of their body with text in the upper part of the picture. In terms of content, the replies in the two verbal conditions were identical (e.g. "*My running shoe is worn-out"*). The two nonverbal (direct and indirect) target pictures showed the character pointing to some relevant situation/object (e.g. *to a worn-out shoe*), or making a "showing" gesture (*holding up a worn-out running shoe*), or a combination of those two (pointing and showing, *pointing to a worn-out shoe which he was holding up for the other to see*). The gesture corresponded to the content expressed verbally in the two verbal conditions. Whether a given target picture was direct or indirect was determined solely by the preceding context, especially the question on the second picture. The only difference between the two direct and indirect target pictures was a subtle change in the speaker's facial expression on some indirect items, to make them situationally appropriate, just as there would be differences in intonation if the items were presented auditorily. Facial expressions accompanying gestures provide cues to the social intention of the speaker (Clark, 1996).

We also constructed a series of probe sentences presented after each target reply, describing a possible continuation of the interaction. The participants judged whether a given outcome was "normal" or "unusual". For the running example, a *normal outcome* probe sentence was "*Only one of the boys will go for a run*", whereas the *unusual outcome* sentence was "*Both boys are heading out for a run*". We equated the normal and unusual versions on the number of words. Each participant saw only one version of the probe sentence after a given item in each condition. The main purpose of the probe sentences was simply to check whether the participants made the correct inference, and exclude trials on which the incorrect inference was made.

In addition, there were 25 filler items designed to make the materials less predictable. Filler either contained an indirect, but affirmative, reply after a yes/no question (as most experimental items implied a negative answers, e.g. a refusal to go running), or an indirect answer after a wh- question.

No participant saw the same cartoon in more than one condition. Thus, even though there were hundred cartoons per condition, each participant received a selection of 25 direct verbal, 25 indirect verbal, 25 direct gestural and 25 indirect gestural, which amounted to 4 stimuli lists. Four additional stimuli lists were made by reversing the order of items on these lists. We took care to construct the stimuli lists so that the following variables were balanced for each condition within a given stimulus list: the number of words on the first two pictures of each cartoon strip (context), and the number of words and frequency of content words in the reply (for verbal conditions). The full set of items is available upon request from the first author.

2.1 Pretesting the stimuli

In order to check the understanding of the gestures, we conducted a paper-and-pencil test with 20 participants on the gestural versions of the items. The participants were asked to write down, in a single sentence, what the characters on the target picture would say if they had not used a gesture. Each participant saw only one (direct/indirect) version of each item, i.e. we collected 10 ratings per item/condition. We then counted how many replies were expressing the message we intended for the target gestures. For example, in the running-shoe example, an acceptable answer would be "*My shoe has a hole in it*," or "*I don't have a proper running shoe*," or "*I cannot run in this shoe*". An inacceptable answer would be "*I have new shoes*," or "*You can run in my shoes*". If a given cartoon received less than 8 acceptable ratings, it was changed.

We conducted another paper-and-pencil pretest where 116 participants rated, on a scale of 1 to 7 (*entirely direct* to *entirely indirect*), how direct/indirect each target reply was. The verbal and gestural versions were rated separately, not to inflate differences between modalities. Even though the difference between direct and indirect replies within each modality was highly significant, in order to maximize the contrast, we changed replies in which the difference was less than 2 points, and re-run these changed items with another 20 participants.

In the final stimulus set, the direct verbal replies were rated on average as 1.97 (0.69), direct gestural as 2.51 (1.02), indirect verbal as 4.83 (0.77) and indirect gestural as 4.23 (0.89), which were highly significant differences (Wilcoxon sign-rank test, verbal $Z = -8.69$, $p < .001$, gestural $Z = -8.3$, $p < .001$). Interestingly, direct gestural replies were rated as less direct than direct verbal replies, and indirect gestural replies were rated as less indirect than indirect verbal replies. We will return to this issue in the Discussion.

2.2. Participants

32 participants were recruited from the Max Planck Institute's participant database and from the Radboud University participant database. All gave their informed consent before participating. Data of 5 participants were excluded due to technical difficulties during data acquisition or excessive movement. The average age of the remaining 27 participants (9 males) was 22 years (range 18 – 32). All participants were right-handed, had normal or

corrected-to normal vision, and did not report any neurological or language deficits. They received financial reward or course credits for their time.

2.3. Experimental procedure

Participants were given instructions, including examples, outside of the scanner. The scanning session consisted of the cartoon task, a structural scan and the two localizer tasks.

The cartoon task contained 100 experimental items and 25 filler items arranged in 10 blocks, with optional breaks in between. Each trial began with an “attention” message presented for 1 second. After that, a jittered fixation cross was presented in the middle of the screen (2.5 – 5.5 s), followed by the two context pictures, which the participants clicked away with a button press after they had finished reading. The target trial was presented after another fixation cross, which was shorter and less jittered (2 – 2.5s in 100ms steps), since we did not want the participants to hold the context stories in their short-term memory for too long. We also did not want to introduce unnaturally long and randomly distributed silent period between the question and reply, as silence can have a communicative function as well. The target frame was presented in the middle of the screen until button-press. After another jittered fixation cross (2 – 3s), the probe statement was presented in the middle of the screen, with the words “normal” and “unusual” in Dutch displayed in the left and right bottom corners. The trial was finished when the participant had pressed the corresponding button with their right index or middle finger. The entire duration of the cartoon task was around 45 minutes.

2.4. ToM localizer

To be able to look for possible modality differences within the mentalizing areas in a more sensitive analysis, we also include a Theory-of-mind localizer task and carry out a region-of-interest (ROI) analysis on the main regions identified there (MPFC, bilateral TPJ, precuneus). The ToM localizer is a Dutch version (M. J. van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012) of a standard localizer task developed by R. Saxe and collaborators (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011), which contains 10 false-belief and 10 false-photograph items delivered as written stories, and contrasts processing of mental vs physical representations. The role of the participant is to read through the stories and then answer a probe question, which requires her to reason about false/true beliefs of the protagonists, or about photographs of physical scenes, which are accurate/inaccurate (i.e. they have or have not physically changed since the photograph was taken).

For example, one false-belief story talked about Sarah getting ready for an evening dance. She placed her shoes under her dress, but her sister later moved them to a location under her bed. The probe sentence said that Sarah gets ready assuming that her shoes are under the dress (*true*). In a false-photograph story, a painter famously painted a picture of a south-bank of a river in 1885. In 1910, a dam was built in the areas and the river basin was flooded, killing the old forests on its banks. Now the entire area is under water. The probe sentence said that on the painting, the south bank is wooded (*true*).

The stories were divided into two blocks of 10. Each story was preceded by a jittered fixation cross (4 – 8s) and lasted for 10 seconds. The probe sentence, during which participants had to make their decision and press the corresponding button, lasted for another 5 seconds.

As some participants did not finish the localizer due to time or technical constraints during their session, we only used data of 23 participants in the final ToM localizer analysis.

2.5. Language localizer

To be able to compare our findings on verbal/gestural indirectness to relatively de-contextualised linguistic processing at the phrase- and sentence-level, we also employ a second localizer task, in which participants have to read written sentences and wordlists.

It consisted of 50 sentences and 50 wordlists, equated for length (9 – 12 words) and presented in alternating blocks. There were two versions of the task, created from a pool of 100 sentences. In each version, half of the sentences was scrambled to create wordlists, and then checked by 3 native Dutch speakers who further scrambled any adjacent words that could potentially form a meaningful phrase. As a result, across participants, each word was equally represented in sentences and wordlists, and wordlists did not contain any meaningful connected phrases. In addition, there were 10 filler lists and 10 filler sentences that contained a non-word. The task of the participants was to read through for comprehension and press a button when they spot a non-word. The stimuli were presented visually, one-word at a time, for 38ms per character, with a jittered fixation cross (1.2 – 2.2s) presented before each sentence/list.

Because some participants did not finish the Language localizer for time or technical constraints during their session, we only used data of 22 participants in the final Language localizer analysis.

2.6. Data acquisition

Participants were scanned in a 3-T Siemens Tim-Trio MRI scanner, using an 8 channel surface coil. We acquired the functional images using a multi-echo EPI sequence (Poser et al., 2006) in an ascending order, with repetition time (TR) 2.35 seconds. Each functional volume consisted of 36 slices of 3 mm thickness, with a 17% slice gap. Voxel size was $3.5x3.5x3mm^3$, with a 224mm field of view. There were 4 echo times in the multi-echo sequence: TE1 = 9.4ms, TE2 = 21.2ms, TE3 = 33ms, and TE4 = 45ms. Flip angle was 90°. To characterize participants' anatomy, a whole-brain T1-weighted MPRAGE sequence was performed, with the following parameters: TR = 2.3s, TE = 3.03ms, 192 slices, voxel size $1mm^3$, FOV = 256.

2.7. Data preprocessing and analysis

We conducted all fMRI data preprocessing and analysis using the Statistical Parametric Mapping software (SPM8, fil.ion.ucl.ac.uk/spm/). The first 30 volumes of each task were used for weight calculation of each of the echoes. The functional echo-planar-imaging BOLD images were then realigned and slice-time corrected, with the resulting functional images co-registered to the participants' anatomical volume, normalized to MNI space and spatially smoothed using a three-dimensional isotropic Gaussian smoothing kernel (FWHM = 8mm). A temporal high-pass filter with a cycle cutoff at 128 s was applied.

In the first-level linear analysis, we modelled the onsets and durations of the four target frames (direct verbal, indirect verbal direct gestural, indirect gestural) from their on-screen presentation until 1.5s after the button-press, as well as the fixation cross baseline for the first fixation preceding the context frame presentation, the context frame, and the task. We also modelled the relevant events for the localizers: onsets and durations of wordlists, sentences, and respective fixation periods for the Language localizer, and one event for the story and task for the false-belief (FB) and false-photograph (FP) conditions, as well as the fixation

cross preceding the stories for the FB localizer. The regressors were convolved with a canonical hemodynamic response function, and realignment parameters were also included in the model, to account for participants' movement. The following images were then defined for each participant, and used in the second-level random effects analysis: direct verbal target > fixation cross, indirect reply target > fixation cross, direct gestural target > fixation cross and indirect gestural target > fixation cross. In the second-level analysis, we used these images in a repeated-measures flexible-factorial model with Modality (verbal, gestural), Indirectness (indirect, direct) and Subjects as factors.
For the FB localizer, we conducted a one-sample t-test FB > FP, and Sentences > Wordlists for the Language localizer.

For the ROI analyses, the mean brain activity in FB and LL ROIs was saved per regressor and per participant, using the SPM toolbox MarsBaR, version .44 (Brett, Anton, Valabreg Random Effects ue, & Poline, 2002). The parameters of all ROIs are presented in the *Results* section.

3. Results

3.1. Behavioral results

A 2x2 repeated-measures ANOVA on reaction time results for the speed of comprehension of target replies was consistent with the results of our previous behavioral study (Bašnáková et al., submitted). We found that indirect replies took a longer time to be processed than direct replies in both modalities, but that this effect was much more pronounced for the gesturally delivered ones. There was a main effect of indirectness (F(1,26) = 16.02, p < .001) and modality (F(1,26) = 7.88, p = .009), but these were qualified by a significant interaction (F(1,26) = 10.33, p = .003). The difference between indirect and direct gestural replies (265ms; highly significant t(26) = 4.32, p < .001) was more than three times the difference between indirect and direct verbal replies (74ms, only trend-level significant, t(26) = 1.85, p = .076). Post-hoc paired-samples t-tests also revealed that the two direct replies were processed at a comparable speed (45ms difference, p = .360), and that indirect gestural replies were processed 236ms slower than indirect verbal replies (t(26) = 3.54, p = $.002_{uncorrected}$, with Bonferroni-corrected p-value for post-hoc comparisons = .0125) (*Figure 2*).

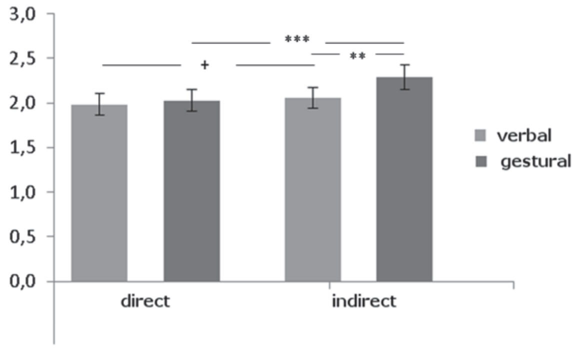


Figure 2. Behavioral results for self-paced reading of verbal and gestural indirectness. Reaction times are in seconds. Error bars denote SEM. Stars indicate significance levels: + $p < .1$, ** $p < .05$, *** $p < .01$.

3.2. Whole-brain analysis results

We were interested in two contrasts in particular: first, the main effect of indirectness (a t-contrast of verbal and gestural indirect against verbal and gestural direct target replies); and secondly, the presence or absence of an interaction between modality and indirectness, showing whether verbal and gestural indirectness engage (partly) different brain regions.

The main effect of indirectness yielded 6 significant clusters. They were located bilaterally in the medial prefrontal and bilateral inferior frontal cortex, including bilateral pars triangularis and pars orbitalis, as well as insular lobe on the left. There was also bilateral angular gyrus (TPJ) activation spanning to superior temporal gyrus on the right side, and supramarginal gyrus on the left. Lastly, there was also a cluster in the right middle temporal gyrus, see *Figure 3* and *Table 2.*

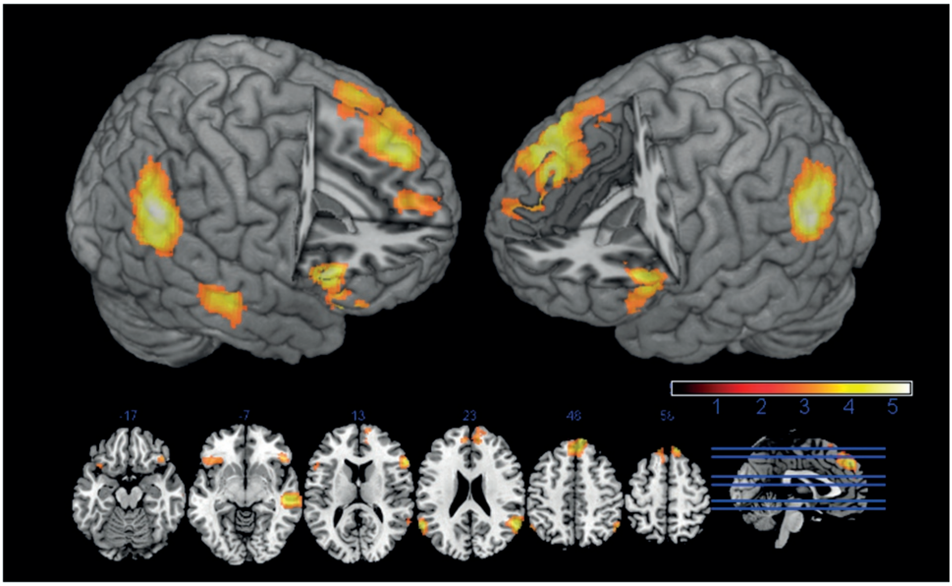


Figure 3. Indirectness effect pooled over the verbal and gestural modality, cluster-level p threshold .005, FWE-corrected. Numbers above slices denote the z-coordinate. Horizontal sections: left is depicted on the left hand-side on this an all subsequent figures.

Anatomical region	Cluster size	T	x	y	Z
A. Main effect of indirectness					
L Superior Medial Gyrus	1998	4.69	-2	42	45
		4.69	2	48	42
		3.76	-4	38	48
		3.64	-6	40	50
R Superior Medial Gyrus		4.57	6	44	46
		4.43	10	46	52
		4.05	10	50	20
		3.78	12	58	18
R Superior frontal Gyrus		3.94	16	52	38
L Posterior-Medial Frontal		3.91	-8	26	66
R Angular Gyrus	1155	5.36	60	-58	34
		4.89	52	-56	32
		4.85	48	-56	34
R Superior Temporal Gyrus		4.37	60	-56	20
L Supramarginal Gyrus	924	5.23	-60	-56	28
L Angular Gyrus		3.72	-48	-60	32
L IFG (p. Triangularis)	672	4.47	-50	22	4
		4.16	-46	24	2
L Insula		3.80	-34	24	0
L IFG (p. Orbitalis)		3.24	-42	24	-8
		3.11	-36	18	-18
		3.06	-50	26	-10
R IFG (p. Triangularis)	636	4.76	56	22	6
R IFG (p. Orbitalis)		4.11	44	28	-14
		3.95	48	28	-12
R Middle Temporal Gyrus	635	4.43	52	-18	-10
		4.41	56	-22	-8
		3.86	66	-22	-8

Table 2. Significantly activated clusters in the Indirectness effect t-test (Indirect > Direct over modalities). Cluster-level p-threshold is .005, FWE-corrected for multiple comparisons. Anatomical labels are based on the Anatomy toolbox in SPM8, MNI coordinates.

In the whole-brain analysis, there were no significant clusters at the .005 threshold for the interaction between modality and indirectness, meaning that there were no differences between verbal and gestural indirectness.

We were not specifically interested in the two simple main effects of modality, since our study was not designed to measure these and our stimulus materials was not identical in this respect (participants read texts in the verbal condition and looked at additional material objects in the gestural condition). The comparison between verbal > gestural replies activated only one cluster located in the left middle temporal gyrus; the opposite contrasts gestural > verbal yielded very large activations in the bilateral temporal, occipital, and parietal cortices, with additional small clusters in the right IFG (pars triangularis) extending into the right middle frontal and bilateral precentral gyri; subcortical regions, including hippocampus and amygdala, were also activated. These two contrasts are characterized in more detail in the *Supplementary materials*, Figure S1 and Table S1.C and S1.D.

3.3. Regions-of-interest analyses

Based on our predictions, we conducted two ROI analyses. We predicted that there should be no modality differences for regions associated with interpreting the speaker's social intention (mentalizing regions), and that there should be a modality x indirectness interaction in the right BA45 involved in referential intention interpretation for non-conventional signals.

a) ROI analysis on mentalizing regions

The false beliefs > false photographs contrast from the ToM localizer yielded a large set of activations which were predominantly temporal. Out of these, we selected ROIs which were consistently activated in ToM studies, according to a metaanalysis by Schurz, Radua, Aichhorn, Richlan, & Perner (2014): bilateral TPJ, MPFC and precuneus. All ROIs were spherical, with 8mm in diameter, with the following coordinates (MNI space): left TPJ (-52 -66 24), right TPJ (58 -62 32), precuneus (left -12 -54 36, right 10 -54 38). For prefrontal activations, we chose two different peaks – one was located more ventrally, in BA10 (-8 60 26), the other more dorsally in BA9 (-10 42 48).

In a 6x2x2 repeated-measures ANOVA with the factors ROI, MODALITY and Random Effects Random Effects INDIRECTNESS, we found a significant main effect of INDIRECTNESS (F (1,26) = 6.33, p = .018), with indirect replies leading to stronger activations than direct replies. However, this effects was qualified by a significant interaction with ROI (F(5,130) = 4.15, p = .002), and a series of paired-samples t-tests showed that the difference between direct and indirect replies was only significant in both TPJs (left: t(26) = 2.68, p = .013, right: t(26) = 3.25, p = .003:) and BA9 (t(26) = 2.37, p = .025); it failed to reach significance in BA10 and precuneus (all p's > .109). Only the right TPJ survived the Bonferroni-corrected alpha level of .01 (.05/5).

As predicted, there was no significant interaction between MODALITY and any of the other factors (*Figure 3A*).

In addition to analyzing ROIs from the ToM localizer task, we also carried out another analysis on mentalizing regions. The reason was that the localizer task exemplifies a fairly specific view on mentalizing; namely, the ability to understand false beliefs. As Schurz et al. (2014) have shown in their metaanalysis, the ToM network is activated by a wide variety of mentalizing tasks, not just false-belief reasoning. From the theoretical standpoint, it is likely that inferences behind detecting the speaker's social intentions are partly different from inferences underlying false-belief reasoning, as is suggested by children who can correctly interpret indirect speech acts at 18 months, i.e. long before they are able to pass a false-belief task (e.g. Schulze & Tomasello, 2015). Therefore, for the second mentalizing-oriented ROI analysis, we selected a more general set of region coordinates, taken from a metaanalysis by Van Overwalle (2009, Fig. 1b; Van Overwalle & Baetens, 2009, Fig. 2): left/right TPJ (MNI -/50 -55 25), MPFC (0 50 20), and precuneus (0 -60 40).

Results from a 4x2x2 (ROI x MODALITY x INDIRECTNESS) repeated-measures ANOVA showed that indirect replies always engaged these areas to a larger extent than direct replies (main effect of INDIRECTNESS, F(1,26) = 13.64, p = .001) at all ROIs and across both modalities. Again, there were no significant two- or three-way interactions of MODALITY with INDIRECTNESS, all p's > .201.

These results suggest that working out the social intention behind indirect replies engages some of the same regions typically activated in mentalizing/socio-cognitive tasks, but that indirectness inferences are not equivalent to false-belief reasoning, as assessed with the false-belief > false-photograph localizer task.

b) ROI analyses on the right IFG

In the second set of ROI analyses, we investigated regions involved in the interpretation of the speaker's referential intention. We hypothesized that the region differentially engaged for verbal and gestural indirectness will be the right BA45, which contributes to interpretation if communicative non-conventional signals (see Introduction). We selected coordinates for the right pars triangularis (BA_T: 50 20 30) based on a study by Tylén, Wallentin and Roepstorff (2009) on material communicative signals. In addition, we also selected a right BA45 ROI from our Language localizer. It was centered around the right-hemisphere equivalent of the most highly activated peak in a left BA45 cluster in the meaningful sentences > wordlists contrast (BA_LL: 48 18 22). There was considerable overlap between these two ROIs.

We had two kinds of control locations: first, we used the left-hemisphere homologues of the two right BA45 ROIs, where we did not predict a specific effect for gestural indirectness (BA45_T -50 20 30, BA45_LL -48 18 22). In addition, we defined a bilateral pars orbitalis (BA47) ROI, which should play a role in indirectness, but not specifically in gestural indirectness. We hypothesized that since bilateral BA47 is involved in situation model building (e.g. Menenti, Petersson, Scheeringa, & Hagoort, 2009), it should be more strongly activated by indirect conditions, but not specifically by gestural indirectness in the same way as the right BA45. Coordinates for BA47 (-/+48 30 -13) were based on Menenti et al. (2009) who studied the interaction between world knowledge violation and local/global context.
We predicted a three-way interaction between LATERALITY (left/right BA45), MODALITY (verbal/gestural) and INDIRECTNESS (direct/indirect); no such effect was predicted for the control location, BA47, even though we expected it to be more active for INDIRECTNESS in general.

We conducted a series of 2x2x2 repeated-measures ANOVAs for each of these ROI pairs, with LATERALITY, MODALITY and INDIRECTNESS as factors. The results were in line with our predictions. For both BA45 ROIs, we found significant three-way interactions (BA45_LL: $F(1,26) = 4.49$, $p = .044$, BA45_T $F(1,26) = 4.47$, $p = .044$). There were also other significant effects, but these were qualified by this interaction: for BA45_T, it was a two-way interaction of LATERALITY and MODALITY ($F(1,26) = 5.92$, $p = .022$), and a main effect of MODALITY ($F(1,26) = 11.06$, $p = .003$); all other p's > .074. For BA45_LL, there was also an interaction of LATERALITY and MODALITY ($F(1,26) = 4.53$, $p = .043$), and a main effect of MODALITY ($F(1,26) = 18.21$, $p < .001$).

To detect the direction of these three-way interaction effects, we carried out post-hoc paired-samples t-tests between verbal and gestural direct replies (direct gestural > direct verbal), and verbal and gestural indirect replies (indirect gestural > indirect verbal). We found that as expected, indirect gestural replies activated the right pars triangularis significantly more than indirect verbal replies ($t(26) = 4.66$, $p < .001$ for BA45_T, $t(26) = 5.20$, $p < .001$ for BA45_LL), and that this was exclusive for the right BA45 (the left BA45 p's for this comparison were > .069). The second largest difference was between direct gestural and direct verbal in the right BA45, with $t(26) = 2.20$, $p = .037$ for BA45_T and $t(26) = 2.73$, $p = .011$ for BA45_LL; although only the latter survived a Bonferroni-corrected p-value threshold of .0125. In the left BA45, the picture was slightly different for the two ROIs: in left BA45_T, it was insignificant ($p = .083$), but for the left BA45_LL, this difference approached significance ($t(26) = 2.44$, $p = .022_{uncorrected}$), with gestural directness activating this area more than verbal directness (see *Figure 3, C and D*).

As for the control condition of BA47, we saw a different pattern of activation. As expected, there was no specific effect for gestural indirectness in the right hemisphere: the three-way interaction between MODALITY, INDIRECTNESS and LATERALITY was nonsignificant ($F(1,26) = 1.47$, $p = .236$). Indirect replies activated this region more than direct replies (main effect of INDIRECTNESS, $F(1,26) = 7.46$, $p = .011$). There was also a trend towards an interaction with LATERALITY ($F(1,26) = 3.72$, $p = .065$), with left BA47 showing a smaller difference between direct and indirect replies than right BA47 (left $t(26) = 2.05$, $p = .051$; right $t(26) = 3.13$, $p = .004$) (*Figure 3, E*).

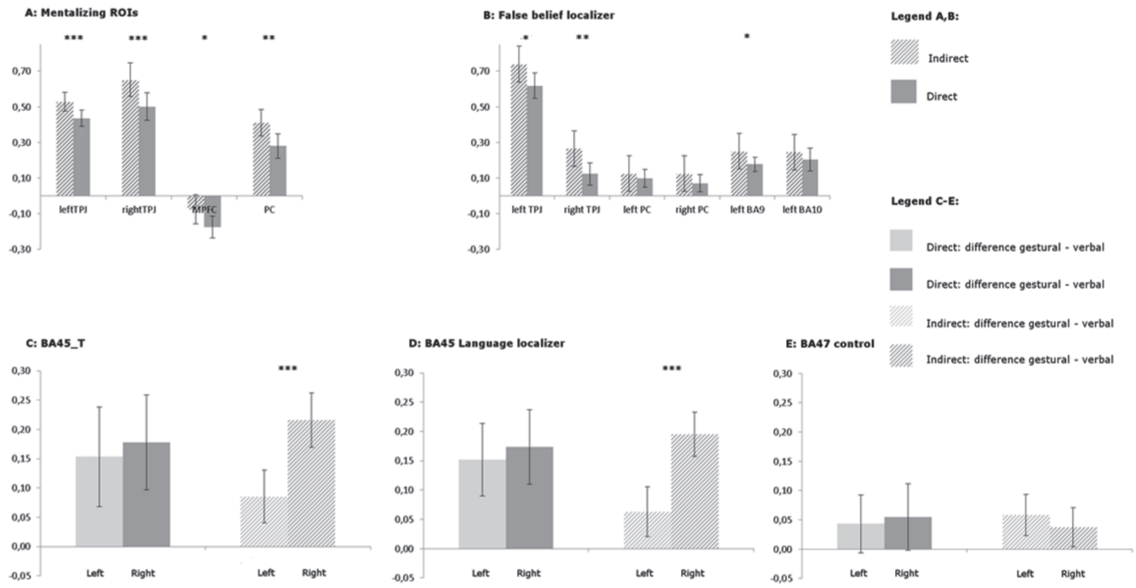


Figure 3. ROI analyses results. A) Indirect and direct replies, relative to a fixation cross baseline, in both modalities in *mentalizing* regions based on ToM metanalysis by van Overvalle (2009); B) Indirect and direct replies in both modalities in *false belief* regions based on False beliefs > False photographs localizer; C and D) Differences between gestural and verbal replies in each modality at the right and left BA45; coordinates for C are based on Tylén et al. (2009), coordinates for D are based on Language localizer; E.) Differences between gestural and verbal replies in each modality at the right and left BA47 (control condition). Stars indicate significance levels: * $p < .1$, ** $p < .05$, ***$p < .01$. Note that the figure displays the findings on difference bar graphs for illustratory purposes, while the analyses presented in the text are based on a comparison of each condition against a baseline.

3.4. Localizers: whole-brain analyses

For completeness, we are also listing significantly activated clusters from the Language and ToM localizers in Table S1 (A and B).

As expected, the ToM localizer engaged bilateral TPJ, MPFC, and precuneus. In addition, there was also a lot of temporal activation, presumably due to the verbal nature of the stories, as we have modelled both the stories and the probe questions as targets, according to the authors' instructions. The language localizer isolated a predominantly left-lateralized network of inferior frontal and temporal regions, typical for language comprehension. In addition, there was also a significantly activated cluster in the right temporal lobe.

Figure 4 shows the overlap between both localizers and the main effect of indirectness. As can be seen, the indirectness effect overlapped to a larger extent with the ToM localizer (TPJ and MPFC), than with the language localizer (mainly temporal regions). The figure also shows that two small areas were activated for all three tasks, located in the left angular gyrus and right middle temporal gyrus.

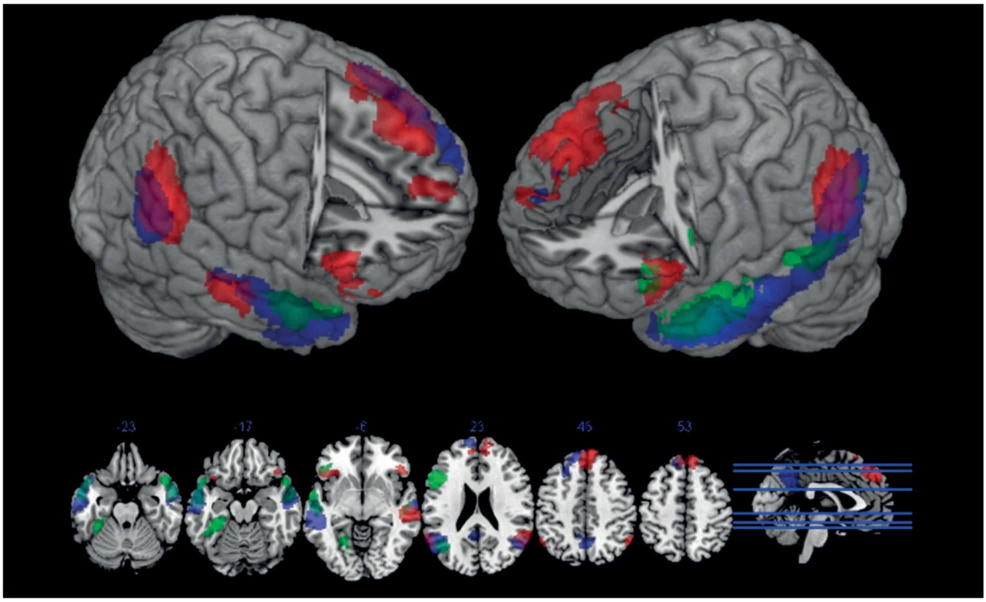


Figure 4. Overlap between the Indirectness effect (red, pooled over both modalities) and both localizers (False beliefs > False photographs in blue, Language localizer in green), cluster-level p threshold .005, FWE-corrected. Numbers above slices denote the z-coordinate.

4. Discussion

This study has two main results. First of all, we have replicated our previous findings on verbal indirectness (Bašnáková et al., 2014, 2015), showing a largely overlapping pattern of activation, even though the present task involved different participants and stimulus material, and the mode of presentation was visual instead of auditory. Secondly, we enriched the existing accounts of the pragmatic inferences, and neural infrastructure supporting them, behind indirect communication in the verbal and gestural modality. We did this by showing that there are no differences between verbal and gestural indirectness in getting at the speaker's social intention (what did she want the recipient to think/do/feel), but that the substantial slowing down in interpreting indirect gestural replies is due to non-mentalistic inferences. Our hypothesis was that the difficulty arises at the level of the speaker's referential intention (what is she directing the recipient's attention to when using verbal utterances or gestures), and we indeed found a differential involvement of the right pars triangularis in gestural indirectness, suggesting that the difficulty lies in the addressee trying to determine what specific aspect of the visual scene is the speaker trying to point out with her gesture. Namely, our results showed that the mentalizing inferences associated with interpreting the speaker's social intention activated the same brain regions regardless of the method of signaling: the core theory-of-mind regions, namely the MPFC, precuneus and bilateral TPJs, were more active for indirect than direct replies independently of whether the speaker used verbal utterances or pointing gestures. In contrast, interpreting the communicator's referential intention was contingent on both the modality of the signal and directness of communication. Pars triangularis (BA45) in the right frontal cortex was significantly more engaged in processing indirect communicative acts expressed via pointing gestures than via speech, suggesting that as communication gets less conventional, and context less constrained, it is harder to zoom in on what the communicative signal picks out, which means specific challenges for interpreting gestural indirectness. One possibility is that interpreting complex verbal and gestural utterances differs in the need to recruit additional contextual cues, possibly in order to resolve a greater degree of non-conventionality, and therefore uncertainty about the referent, associated with the gestural signal. This ambiguity is greatly reduced in the case of verbal indirectness, because language allows the speaker to offer a more nuanced and accurate description of reality. In the following, we will discuss these findings in more detail, as well as limitations and implications of the current study.

4.1. Interpreting speakers' social intentions: no difference between modalities

One of the main aims of this study was to characterize the pragmatic inferences behind indirect meaning comprehension in both verbal and gestural domains. Based on theoretical accounts (Tomasello, 2008; Buchiarelli et al., 2006), we expected that the intentional-inferential infrastructure underlying the interpretation of the speaker's social intention – what she really meant - will be identical across modalities, and we have confirmed this expectation. There was no interaction between verbal and gestural indirectness within the mentalizing regions neither in the whole brain analysis, nor in the more sensitive ROI analysis. Thus, it seems that bridging the gap between the speaker's referential and social intention is indeed independent from the means of expression the speaker uses.

The involvement of an identical intentional network, the IPN, has been shown in a previous study comparing direct communicative verbal utterances and gestures (Enrici et al., 2011). However, there were important differences between our studies. Firstly, Enrici's study focused on direct communicative acts in which the gap between the speakers' social and referential intentions was minimal: in Enrici's study, what the speaker said (or pointed/gestured to) was largely equivalent to what she meant. Secondly, by comparing these acts to a baseline which was entirely non-communicative (and agent-less), the recipients had to infer not only the speaker's referential and social intentions, but also that she had a communicative intention at all. Our study has a more refined focus, zooming in on the difference between referential and social intentions, with utterances/gestures always embedded within communicative situations. This is also a likely explanation for the differences in activation between our indirectness effect and the IPN, which lacks any regions associated with discourse-level processing, such as IFG or middle temporal areas.

Due to using a ToM localizer comparing false beliefs with false photographs that we administered to our participants, we can also be more specific on the nature of the mentalizing inferences behind speaker's social intention interpretation. Figure 4 shows that the overlap between the indirectness effect and false-belief processing was only partial; this was also confirmed in a ROI analysis on representative locations from the localizer (bilateral TPJ, precuneus and 2 peaks in the MPFC), out of which only the right TPJ was reliably activated by indirectness to a larger extent than by directness. There was a trend in the correct direction for the left TPJ as well, but these activations were not strong enough to survive the corrected threshold. In contrast, there was reliable overlap of indirectness with slightly different locations within the same regions (TPJ, MPFC, precuneus) taken from a metaanalysis of more general mentalizing tasks.

One tentative explanation of this discrepancy refers to differences between inferential processes involved in false-belief reasoning and indirect communication. First of all, the false-belief > false-photograph localizer requires the participants to *explicitly* reason about others' false beliefs, and it is unlikely that this level of explicitness is routinely employed in conversations.

An interesting point from the pragmatic literature is that indirect meanings come in a great variety and in some contexts can be interpreted even without references to the speaker's intentions. Kissine (2015), based on the work of Jary (2013), offers an example of so-called material implicatures. If somebody is in a room with an open window, and utters "It's really cold in here", the recipient can take this as an indirect request implying that she should close the window. However, in some instances, it is possible to understand the implied portion of the message (*close the window*) purely based on what the recipient knows about the world – e.g. that in cold seasons, closing windows makes the air warmer. Certainly, this simplified reasoning would not be possible for face-saving type of indirectness ("*It's hard to give a good presentation*,") or for verbal irony, but it is conceivable for some of the more simple indirect utterances which were used in our stimulus material. The resulting picture, therefore, is one of a gradient of inferential difficulty for indirect utterances, with some presumably engaging a more limited range of mentalizing areas, to a possibly full-blown, explicit false-belief reasoning in contextually less supported situations, where the recipient has to deal not only with interpreting the information encoded in the utterance, but also with e.g. the social consequences of providing or withholding sensitive information. Importantly, however, the results of our study suggest that whatever is the true identity of the mentalizing inferences, the modality of the signal should not play a role. For the recipient, any ostensive signal, be it an elaborate verbal utterance or a simple pointing gesture, will trigger a search for the social intention of the speaker.

4.2. Interpreting speakers' referential intention: difference between modalities

The most novel aspect of our findings concerns the interpretation of gestural indirectness. We have found that the right BA45 in the inferior frontal gyrus was significantly more active for gestural than for verbal indirectness; no such difference existed in the control location in the left hemisphere. Our interpretation of this effect is that this region is likely involved in the inferences behind establishing the speaker's referential intention, especially from less conventional signs such as pointing gestures (see also Tylén et al. 2009 on interpreting social meaning of material objects). This is consistent with what we hypothesized in our previous behavioral study (Bašnáková et al., submitted): that recipients struggle with establishing the referential intention behind indirectly used pointing gestures to a much larger extent than for indirect verbal utterances, and that this is what makes gestural indirectness more difficult to interpret.

We suggest that establishing what an indirect pointing gesture "picks out" in the speaker's environment is more difficult than that of a verbal utterance, due to two factors. One of them is that the gesture's interpretation always depends on the context. Secondly, as this context gets less constrained, the uncertainty of what the gesture refers to increases as well, determining not only what part of the broadly indicated visual scene is relevant, but also how is it relevant to what the speaker actually wants to communicate.

In theory, this is not limited to the gestural modality itself. As a comment by Clark (1997) on one of Grice's (1975, 2008) own examples shows, even the referential meaning of *verbal* utterances is contingent on the utterance's implicated meaning. In this example, speaker A says "*I'm out of petrol,*" and speaker B replies "*There's a garage around the corner*". Since "garage" has at least two meanings in British English, both related to cars (parking structure/gas station), the listener has to determine the appropriate meaning of this expression in order to be able to interpret the utterance; however, this would be impossible without determining what B means (his social intention), i.e. that A can get petrol there.

In general, language is without a doubt in a better position to describe complex reality in a less ambiguous way than pointing gestures, and it typically does. However, it might be that the "referential intention ambiguity effect" that we have found here for gesture, is not necessarily gesture-specific. In theory, recipients might struggle to establish a clear referential intention even from language, e.g. when the speaker's verbal expression is very ambiguous and needs considerable contextual support for disambiguating – such as when it contains words with multiple meanings which cannot be straightforwardly interpreted in one way or the other. Our study was not designed to test this prediction, but future studies can contrast more and less ambiguous verbal expressions in indirect communication (see e.g. Nieuwland, Otten, & Van Berkum, 2007).

We have also shown that while the referential intention ambiguity effect between the verbal and gestural modality in the right pars triangularis is largest for the indirect stimuli, there is also a trend towards a corresponding difference (gestural > verbal) for the direct stimuli. This is not inconsistent with our predictions, because indirectness is not an "all or nothing" phenomenon, but rather exists on a continuum. As explained in the Introduction, it is reasonable to expect that a referent of a direct pointing gesture will still require some contextual support to resolve what exactly it "picks out" in the environment, although to a smaller extent than indirect pointing gestures.

As for the control conditions, the right BA47 showed only a general effect for indirectness. This corresponds to a view that bilateral BA47 is involved in situation model updating, which should be more demanding for indirect utterances (see e.g. Bašnáková et al., 2014, 2015, Menenti, 2007). We did not expect to find a modality difference here, since once the inferences have been made (e.g. that it is impossible to run in worn-out shoes), they should be equally difficult to integrate into the recipient's situation model regardless of whether the original signal was verbal or gestural. In other words, the result of the inference, which is used to update one's mental model of the unfolding conversation, does not depend on the means which were used to produce an ostensive signal.

Lastly, our results speak to the existing theories about communicative acts in different modalities. Rather than going against Tomasello's (2008) and Bara's (2010) accounts about verbal and gestural indirectness being equivalent in terms on the pragmatic inferences they require, we can rather think of our results as complementary to these theories. If we only consider the social intention of the speaker, our findings are in complete agreement, as well as with the conclusions of the studies testing them (Bucciarelli et al., 2003, Cutica, 2005, Schulze & Tomasello, 2015): the inferential infrastructure underlying social intention interpretation is identical across modalities. What differs, are non-mentalizing inferences about the speaker's referential intentions. This particular aspect has not been foregrounded in any of these theories, and none of their empirical studies is sensitive enough to detect potential differences at the level of the speaker's referential intention. However, Cutica's (2005) results are suggestive of such differences, as she found both RHD patients and healthy controls struggle more with indirect gesture than indirect speech. Our fMRI and behavioral study have a much larger number of experimental items in each condition, and our measurement methods are better suited to detect more subtle differences between conditions than measuring error frequencies, which was the method used in previous studies.

4.3. Additional considerations

One interesting question is whether we can generalize our gestural indirectness findings to any types of gestures, or whether they are specific to pointing. As we have already argued previously in the Discussion, it is likely that the ambiguity associated with the speaker's referential intention in indirect communication is not necessarily confined to the gestural modality. Even verbal indirect utterances can be ambiguous to the point that their literal meaning is hard to establish, although it is generally less likely than with pointing. We hypothesize that while referential meaning ambiguity can be a problem in any kinds of communicative signals used indirectly, the severity of his issue might depend on how conventional the signal is. Pointing gestures represent one end of a spectrum of communicative signals which have no inherent meaning, and which completely depend on the context for disambiguation. Conventionalized gestures like emblems (*thumbs up*) already do, on the other hand, contain a lot of information in the signal itself, even though they are processed differently from linguistic expressions (e.g. Husain, Patkin, Kim, Braun, & Horwitz, 2012). Linguistic expressions have a conventional meaning, but can still be used in a way that the recipient has a hard time selecting the contextually appropriate one, or can be even used non-canonically, such as vernacular language. Therefore, we would expect that in theory, recipients can have problems with any types of signals used in indirect communicative settings, but typically, there would likely be a continuum from pointing gestures, through emblems, to verbal utterances. To our knowledge, the nature of the neural correlates of such different types of ambiguity in the referential intention has not yet been investigated within a single study.

Finally, we would like to address the issue of the ordering of pragmatic processes. Referring to different levels of pragmatic meaning analysis (referential and social intention, or what is said and what is meant) might imply that the analysis of what is said/referential intention has to be completed before the recipient can move to considering what is meant/social intention. Our fMRI data cannot directly speak to the temporal ordering of different processes, but based on both theoretical considerations, as well as research in the neurobiology of language (e.g. Hagoort & Indefrey, 2014), such strict temporal ordering is unlikely. Utterance-level (referential) and implicature-level (social) meaning are likely derived interactively, since it is often impossible to determine the specific meaning of words (or gestures) without working out, at least in part, the implicated meaning of the entire utterance (Clark, 1997). Gestural indirectness is a good example of this – in order to determine what specific part of the shoe picked out by the pointing gesture is relevant in the running situation, the recipient must at least begin to understand (with the help of contextual information, including facial expressions) that the speaker is declining his invitation for a run.

4.4. Limitations and future directions

This study contributes to our understanding of the neural underpinnings of communication by including types of communicative situations which have not been extensively studied before: pointing gestures used to express more complex, indirect messages. One of the strong points of the design is its well-matched conditions, where the target pictures/text in each modality are identical. The downside of this approach is that the communicative acts are presented in relative simple cartoons, where such tight experimental control is achievable.

There is a growing realization in the neurobiology of language literature that the field needs to move toward more ecologically valid paradigms, especially when studying communicative situations (e.g. Bara, Enrici, & Adenzato, 2015; García & Ibáñez, 2014). In this context, one of the potential improvements would be to make the design less situationally constrained and more interactive. With our approach, we achieved that there was a reasonable level of uniformity in the amount and nature of contextual information all participants had at their disposal; at the same time, in real-life conversation, it is exactly the unconstrained nature of contextual information that makes inferencing difficult. There are potentially myriads of information present – tone of voice, social motives, status differences, facial expressions, competing personal goals, etc. - and in order to lead a successful dialogue, the recipient must be able to identify and zoom in on the most relevant ones at the given moment. Therefore, we suggest that future studies embed gestural communication in situations that are sufficiently rich: for example, by using multimodal stimuli (e.g. videos) or actors, and also by studying situations in which the use of gesture is pragmatically motivated, e.g. the speaker wants to hide certain communication from somebody else's view.

It would be also interesting to study the *development* of indirect gestural communication at the neural level. We know that infants use pointing gestures earlier than they communicate verbally, and that they are able to understand ostensive indirect communicative acts early on, by about 18 months of age (Schulze & Tomasello, 2015). Studying the development of indirect gesture comprehension would give us the opportunity to gain access to the way brain deals with contextual information in communication, which is a skill that goes well beyond verbal code processing.

5. Conclusion

By going beyond the verbal code, and investigating the similarities and differences behind indirect verbal utterances and communicative gestures at the brain level, we have contributed to an ever more detailed characterization of the neural infrastructure of interpersonal communication. We have presented evidence suggesting that there are several levels of intentions involved in indirect gestural communication; in the future, the referential-social intention distinction can potentially contribute by adding another layer of detail into research on the development and impairments of pragmatic communicative abilities.

References

Ackeren, M. Van, Casasanto, D., & Bekkering, H. (2012). Pragmatics in action: indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive*.

Bach, K. (1994). Conversational Impliciture. *Mind & Language*, *9*(2), 124–162. https://doi.org/10.1111/j.1468-0017.1994.tb00220.x

Bara, B. G., Enrici, I., & Adenzato, M. (2016). Chapter 54 - At the Core of Pragmatics: The Neural Substrates of Communicative Intentions. In *Neurobiology of Language* (pp. 675–685). https://doi.org/http://dx.doi.org/10.1016/B978-0-12-407794-2.00054-7

Bašnáková, J., van Berkum, J., Campisi, Mentel, A., & Hagoort, P. (n.d.). Did you get my point? How recipients understand indirect communicative gestures.

Bašnáková, J., van Berkum, J., Weber, K., & Hagoort, P. (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers? *Neuropsychologia*, *76*, 79–91. https://doi.org/10.1016/j.neuropsychologia.2015.03.030

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the language given: the neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578. https://doi.org/10.1093/cercor/bht112

Bezuidenhout, A. (2005). Review: Thoughts and Utterances: The Pragmatics of Explicit Communication. *Mind*, *114*(455), 722–728. https://doi.org/10.1093/mind/fzi722

Brett, M., Anton, J.-L. L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. In *NeuroImage* (Vol. 16, p. abstract 497).

Brown, P., & Levinson, S. C. (1987). *Politeness: some universals in language usage*. Cambridge University Press.

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., … Aureli, T. (2014). Do You Know What I Mean? Brain Oscillations and the Understanding of Communicative Intentions. *Frontiers in Human Neuroscience*, *8*, 36. https://doi.org/10.3389/fnhum.2014.00036

Bucciarelli, M., Colle, L., & Bara, B. G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, *35*(2), 207–241. https://doi.org/10.1016/S0378-2166(02)00099-1

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*(3), 567–598. https://doi.org/10.1080/01638539709545003

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, *58*(2), 332–373.

https://doi.org/10.1353/lan.1982.0042

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In *Elements of discourse understanding* (pp. 10–63).

Committeri, G., Cirillo, S., Costantini, M., Galati, G., Romani, G. L., & Aureli, T. (2015). Brain activity modulation during the production of imperative and declarative pointing. *NeuroImage*, *109*, 449–457. https://doi.org/10.1016/j.neuroimage.2014.12.064

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage*, *55*(2), 705–712. https://doi.org/10.1016/j.neuroimage.2010.12.040

Enrici, I., Adenzato, M., Cappa, S., Bara, B. G., & Tettamanti, M. (2011). Intention Processing in Communication: A Common Brain Network for Language and Gestures. *Journal of Cognitive Neuroscience*, *23*(9), 2415–2431. https://doi.org/10.1162/jocn.2010.21594

Feng, W., Wu, Y., Jan, C., Yu, H., Jiang, X., & Zhou, X. (2017). Effects of contextual relevance on pragmatic inference during conversation: An fMRI study. *Brain and Language*, *171*, 52–61. https://doi.org/10.1016/j.bandl.2017.04.005

García, A. M., & Ibáñez, A. (2014). Two-person neuroscience and naturalistic social communication: the role of language and linguistic variables in brain-coupling research. *Frontiers in Psychiatry*, *5*, 124. https://doi.org/10.3389/fpsyt.2014.00124

Garrett, M., & Harnish, R. M. (2007). Experimental pragmatics: Testing for implicitures. *Pragmatics & Cognition*, *15*(1), 65–90. https://doi.org/10.1075/pc.15.1.07gar

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax And Semantics* (Vol. 3, pp. 41–58). Academic Press.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Grice, H. P. (2008). Further notes on logic and conversation. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of Human Inference and Its Foundations*. Cambridge University Press.

Hagoort, P., & Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, *37*(1), 347–362. https://doi.org/10.1146/annurev-neuro-071013-013847

Holtgraves, T. (1998). Interpreting indirect replies. *Cognitive Psychology*, *37*(1), 1–27.

Husain, F. T., Patkin, D. J., Kim, J., Braun, A. R., & Horwitz, B. (2012). Dissociating neural correlates of meaningful emblems from meaningless gestures in deaf signers and hearing non-signers. *Brain Research*, *1478*, 24–35. https://doi.org/10.1016/j.brainres.2012.08.029

Jang, G., Yoon, S. A., Lee, S. E., Park, H., Kim, J., Ko, J. H., & Park, H. J. (2013). Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. *NeuroImage*, *81*, 61–72.

Jary, M. (2013). Two types of implicature: Material and behavioural. *Mind and Language*, *28*(5), 638–660. https://doi.org/10.1111/mila.12037

Kissine, M. (2015). Pragmatics as Metacognitive Control. *Frontiers in Psychology*, *6*, 2057. https://doi.org/10.3389/fpsyg.2015.02057

Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (2009). When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience*, *21*(12), 2358–68. https://doi.org/10.1162/jocn.2008.21163

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. A. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(2), 228–236.

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(3), 833–838. https://doi.org//10.1073/pnas.0707192105

Recanati, F. (2004). *Literal meaning*. Cambridge University Press.

Redcay, E., Velnoskey, K. R., & Rowe, M. L. (2016). Perceived communicative intent in gesture and language modulates the superior temporal sulcus. *Human Brain Mapping*, *37*(10), 3444–3461. https://doi.org/10.1002/hbm.23251

Shibata, M., Abe, J. ichi, Itoh, H., Shimada, K., & Umeda, S. (2011). Neural processing associated with comprehension of an indirect reply during a scenario reading task. *Neuropsychologia*, *49*(13), 3542–3550. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2011.09.006

Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition*, *136*, 91–98. https://doi.org/10.1016/j.cognition.2014.11.036

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Sperber, D., & Wilson, D. (1995). Relevance: communication and cognition. *Behavioral and Brain Sciences*, *2nd*(4), 697–710. https://doi.org/10.1017/S0140525X00055345

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Tylén, K., Wallentin, M., & Roepstorff, A. (2009). Say it with flowers! An fMRI study of object mediated communication. *Brain and Language*, *108*(3), 159–166. https://doi.org/10.1016/j.bandl.2008.07.002

Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a Tool for Interacting Minds. *Mind & Language*, *25*(1), 3–29. https://doi.org/10.1111/j.1468-0017.2009.01379.x van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S.-A. (2012). Pragmatics in Action: Indirect Requests Engage Theory of Mind Areas and the Cortical Motor Network. *Journal of Cognitive Neuroscience*, *24*(11), 2237–2247. https://doi.org/https://doi.org/10.1162/jocn_a_00274

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, *30*(3), 829–58. https://doi.org/10.1002/hbm.20547

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*(3), 564–584. https://doi.org/https://doi.org/10.1016/j.neuroimage.2009.06.009

Walker, T., Drew, P., & Local, J. (2011). Responding indirectly. *Journal of Pragmatics*, *43*(9), 2434–2451. https://doi.org/10.1016/j.pragma.2011.02.012

Wilson, D., & Sperber, D. (1986). Pragmatics and modularity. *In: Farley, A and Farley, T and McCullough, K, (Eds.) (Proceedings) Chicago Linguistic Society 22. (Pp. Pp. 67-84). Chicago Linguistic Society: Chicago, IL. (1986)* .

Yang, J., Andric, M., & Mathew, M. M. (2015). The neural basis of hand gesture comprehension: A meta-analysis of functional magnetic resonance imaging studies. *Neuroscience & Biobehavioral Reviews*, *57*, 88–104. https://doi.org/10.1016/j.neubiorev.2015.08.006

Acknowledgments

JB was partly supported by VEGA grant 2/0085/17; this research was also supported by NWO Vici grant #277-89-001 to JvB.

CHAPTER 6: SUMMARY & DISCUSSION

1. Summary of the results

In this thesis, I investigated the neural correlates of interpreting particularized conversational implicatures triggered by indirect replies. Indirect replies provide an interesting window into communicative language processing for several reasons. First, just like in many other cases in ordinary language use, they convey more than is "in" the words and sentence structure alone, and thus provide suitable testing grounds for studying the inferences that the addressee needs to make to get from utterance meaning to speaker meaning. Indirect replies operate on the interface of the linguistic and the social, since indirectness is often motivated by interpersonal considerations. Secondly, the interpretation of indirect replies places a much heavier emphasis on contextual integration – as indirect replies change their meaning depending on the context. And lastly, the interpretation of indirect replies requires the addressee to reason about the speaker's social intentions – and thus getting at a level of representation that is not purely linguistic, but that is crucial for successful communication.

In the four studies presented in this thesis, I used functional magnetic resonance imaging and reaction time measurements to look at indirectness interpretation from various angles. In Chapter 2, I contrasted indirect face-saving and "information-economy" replies to matched direct replies. The comparison yielded three sets of activations beyond the core language network. First, there were the medial prefrontal cortex (mPFC) and the right temporo-parietal junction (TPJ) activation, both considered the "core" mentalizing areas. Secondly, interpreting indirectness activated several right-hemisphere homologues of language areas: most notably, those associated with contextual integration, such as the inferior frontal gyrus (IFG) and middle temporal gyrus. Third, there were also activations associated with affective processing, including the anterior cingulate cortex (ACC) and insula. We refer to this pattern of activations in three sets of brain regions associated with mentalizing, discourse-level integration, and emotional processing, as the "indirectness effect".

Interestingly, contrasting face-saving indirect replies did not lead solely to affective activations, but also regions associated with contextual integration (right inferior frontal gyrus, BA47). This suggests that addressees might use the knowledge of rules (maxims) governing conversations as one of the contextual cues guiding interpretation. In other words, knowing that this is a "politeness" context might help justify the seemingly irrelevant reply, which in turn triggers inferences about what the speaker actually meant to convey with her utterance.

One of the basic questions motivating this study was whether interpreting indirectness requires inferences about the speaker's intentions. In showing that the indirectness effect involves activations in brain regions associated with theory-of-mind processing, my results support intentional accounts of communication. At the same time, they suggest that simple "simulation" accounts of language comprehension are not plausible models of speaker meaning interpretation.

While the direct and indirect replies were delivered within auditorily presented dialogs, and thus in a fairly natural way in Chapter 2, they were still addressed to other people, and were about other people, which made them not directly emotionally relevant to the participants. This is not typical of how we use language in everyday situations, where we are usually direct addressees. Thus, indirect face-saving remarks are designed by speakers to influence the addressees' perception of them. Manipulating personal relevance of indirect replies was the aim of the study reported in Chapter 3. In a mock job interview, participants sometimes acted as direct addressees, and sometimes as overhearers, of direct and indirect replies of several alleged job candidates. Indirect replies were used in situations when the candidates wanted to cover up lack of skills or attempted to make themselves look better. There were two main results: first of all, I replicated the indirectness effect (indirect > direct replies) from Chapter 2, with a different group of participants and a different set of materials. Secondly, it turned out that - at least in these minimally interactive conditions – interpreting indirectness when the participant *is* the addressee does not significantly change activations in regions associated with mentalizing or contextual integration. The only significant difference was in regions implicated in emotional salience, possibly reflecting the fact that an indirect reply addressed to a participant is meant to manipulate *his/her* feelings and impressions.

Lastly, Chapters 4 and 5 present studies that explore indirectness across modalities by contrasting verbal indirectness with non-verbal indirect replies. In a reaction time (Chapter 4) and an fMRI study (Chapter 5), I compared indirect replies expressed via a verbal utterance or a pointing or showing gesture. I expected that interpreting the speaker's social intention behind her indirect reply should be independent of the modality of the communicative signal. The behavioral results suggested that it is neither the gestural modality, nor indirectness, that makes interpretation more difficult. Rather, it is a combination of the two: when the signal gets more ambiguous, and therefore harder to interpret at its utterance level, it is also more taxing to recover the speaker's meaning, as these two levels clearly interact. The results of the fMRI study seem to suggest that mentalizing networks are equally active for both verbal and gestural indirectness. Based on the results of the behavioral study, I also conducted a region-of-interest (ROI) analysis on the right BA45, involved in processing semantic ambiguity associated with non-conventional signals (Tylén, Wallentin, & Roepstorff, 2009). The analysis revealed that there was higher activation for gestural than for verbal indirectness, which I interpreted as giving support to the conclusion of the behavioral study. Importantly, I again replicated the general indirectness effect, even though the stimuli were now presented visually and not auditorily as in the previous studies.

In the following, I will discuss the implications and limitations of these findings, some methodological considerations, as well as future directions for indirectness research.

2. Pragmatic inferences behind indirectness

The main goal of this thesis was to characterize the inferences behind speaker meaning interpretation. I have shown that getting at the intended meaning of an indirect reply requires several kinds of inferences; most importantly, those implicated in thinking about the beliefs and intentions of the speaker. In this section, I will discuss the findings in more detail.

Based on Grice's and Relevance theory's (RT) analyses presented in the Introduction, successful speaker meaning recognition requires that the listener detects the speaker's social (or, in RT terms, communicative and informative) intention. In other words, communication is essentially intention-recognition, not simple decoding. For this reason, I expected to find regions involved in mentalizing/ToM inferences in the subtraction of indirect > direct utterances. This was supported by the data, as in all three fMRI studies, the core ToM regions – the medial prefrontal cortex and the right TPJ – were part of the indirectness effect.

What exactly do these activations mean? In every chapter, I have talked about pragmatic inferences, mentalizing inferences, ToM-like inferences, and so forth. One of the goals of this thesis was to be more exact about what kinds of inferences support indirectness interpretation. At present, the exact functional role of each of the brain regions involved in ToM processing is far from clear. However, based on a metaanalysis by Schurz et al. (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014), mPFC and right TPJ are the two core areas involved in most ToM tasks. These researchers have identified several task profiles for mentalizing studies, with the best fit for recovering speaker's intentions being "rational actions". Molenberghs and colleagues (Molenberghs, Johnson, Henry, & Mattingley, 2016) have made an important distinction between two systems that subserve ToM: an explicit one and an implicit one. It is likely that intention recognition in communication usually falls under the scope of the implicit one. This suggests that addressee do not typically have to explicitly reason about the speaker's mental states, in contrast to what is presumably going on in tasks like the "false beliefs" one.

This assumption is further supported by the results of Chapter 5, where I use a theory-of-mind localizer that is contrasting a false belief and a baseline false photograph task, and thus can make a direct comparison between processing related to indirectness and to false beliefs. Interestingly, these are not entirely overlapping. Activations are in the same general regions, but peak coordinates differ. This is direct evidence that intention recognition subserving communication is not identical to the effortful, explicit reasoning about false beliefs.

A related question is whether all indirectness always requires intention recognition. Kissine (2015) suggests that a psychologically plausible model should make a distinction between accessibility-based and ToM-based pragmatic processes. For example, it is plausible that some inferences can be made purely on the basis of world knowledge – that a pregnant woman cannot drink wine (and therefore her utterance "*I am pregnant*" after offering her a glass of wine automatically means that she does not want it), or that one cannot run in worn-out running shoes (and therefore a pointing gesture towards them signals that he cannot go for a run). The results reported in Chapter 5 do not seem to fully support Kissine's claim, as there is mPFC and TPJ activation even for informative indirect replies, which are exactly the kinds of utterances Kissine refers to. On the other hand, I cannot rule out the possibility that a more fine-grained analysis would show that *some* of the stimulus items would be interpreted without reference to intentions (as the activations are based on aggregated results across all items).

Apart from mentalizing, it is likely that there are additional cognitive functions implicated in indirectness interpretation, for example executive functions subserved by the prefrontal cortex. There is evidence from clinical populations suggesting that prefrontal dysfunction affects processing of indirect remarks. For example, McNamara and Holtgraves (Holtgraves & McNamara, 2010b; McNamara, Holtgraves, Durso, & Harris, 2010) examined the comprehension of indirect replies in patients with striatal-prefrontal dopaminergic networks damage caused by Parkinson's disease (PD). PD patients were slower than control participants to comprehend the indirect, but not direct, meaning of utterances. This was correlated with the degree of impairment in striatal-prefrontal network, measured indirectly in a task on executive functions (Stroop). In another study, these authors showed that PD patients experience problems in producing polite indirect remarks (Holtgraves & McNamara, 2010a). One reason for these difficulties could be that executive functions might be important for suppressing the meaning that is explicitly coded (ibid.). Alternatively, impaired executive functions might make it harder for people to accurately identify social status differences among interactants and to follow the tacit rules of communicative interactions, which then makes interpretation of face-saving indirect remarks difficult.

Another robust finding of this thesis is that the majority of activations for indirect replies were in the right hemisphere. This is in contrast to the findings of two meta-analyses on non-literal meaning interpretation reported in the Introduction (Bohrn, Altmann, & Jacobs, 2012; Rapp, Mutschler, & Erb, 2012), where most activations, both in terms of peak number and ratio, were in the left hemisphere. However, as I have already foreshadowed in the Introduction, the main reason for this can be that most of the non-literal language in the two metaanalyses were GCIs, which did not need as extensive use of contextual sources as do PCIs investigated in this thesis. In line with the coarse semantic coding theory of RH involvement, the more must listeners rely on contextual information for interpretation, the greater is the RH involvement (Beeman, 1998). For example, in a study by St George, Kutas, Martinez, & Sereno (1999), when participants read paragraphs which did not have a title and therefore it was difficult to establish their coherent situation model, there was more right-hemisphere temporal activation than in the (titled) control condition. Similarily, Menenti (Menenti, Petersson, Scheeringa, & Hagoort, 2009) showed that the right hemisphere was activated by highly context-dependent inferences that were necessary for establishing a coherent situation model for stories that required drawing on contextual information.

The core regions for processing indirectness are also partly overlapping with what Bara et al. call "the intention processing network" (IPN) (Bara, Enrici, & Adenzato, 2016). This network, consisting of the medial prefrontal cortex, precuneus, bilateral superior temporal sulcus and TPJ, is based on studies comparing verbal or nonverbal communicative intentions (*somebody pointing to a bottle to request it*) with a baseline of non-intentional physical causality (*a shelf* *falling off the wall*). Even though this is superficially similar to the contrasts in Chapter 5 (gestural > verbal communicative acts), the IPN has a conceptually different focus. In Chapter 5, both indirect and baseline direct replies are communicative, whereas in the studies reported by Bara et al., the baseline conditions are non-communicative (in fact, they are also agent-less, since they do not involve any actors, just objects and physical actions on them). Therefore, one can think about indirectness regions to be a subset of IPN.

Finally, there are several regions in the indirectness fMRI effect which are also frequently seen in studies on prosody processing. Brain regions involved in prosody processing are the superior and middle temporal gyrus, inferior frontal and orbitofrontal gyrus, as well as insula and temporo-parietal junction (Frühholz, Ceravolo, & Grandjean, 2012; Schirmer & Kotz, 2006). An interesting question is to what extent prosody contributes to the recognition of speaker meaning. In several theoretical accounts, affective prosody is the main determinant of stance (Tomasello, 2008; Van Berkum, 2015). There is also evidence that listeners can detect simple speech acts (= speaker's intentions) from single-word utterances based solely on their prosody, without any semantic information (Hellbernd & Sammler, 2016). However, outside of constrained experimental conditions, prosody is only one of the cues that contribute to the recognition of the speaker's intention. Interestingly, some of the regions involved in prosody processing were also active in the study in Chapter 5, where replies were delivered in the visual modality, as texts. A tentative explanation is that even when participants are just trying to read texts, they might attempt to add prosody which can help them with interpretation.

3. Other studies on indirectness

During the completion of this thesis, several other authors have investigated indirectness as well. What are the commonalities and differences to my findings, and what implications do they have?

Jang et al. (2013) investigated conversational implicatures in question-answer pairs where the question remained the same (*"Is doctor Smith in his office now?"*), but the answer differed in terms of how explicitly it addressed the question. Apart from the explicit condition (*"Dr. Smith is in his office now"*), there were two implicit conditions: a moderately implicit one (*"Dr.*

Smith's car is parked outside the building") and a highly implicit one (*"The black car is parked outside the building"*). The main difference between the moderate and highly implicit conditions was that the latter lacked coherence markers (*"Dr. Smith"*) and the listener thus had to fill in implicated meanings to establish this coherence. Question-answer pairs in this study were presented visually, without any previous context, as if uttered by a pair of anonymous speakers. Regions identified in the highly implicit > explicit contrast, which is the closest one to my indirect > direct comparisons, were similar to the studies reported here. One interesting difference is that they were predominantly left-lateralized, as opposed to my activations which were mostly bilateral or right-lateralized. There might be several reasons for this discrepancy, including the fact that the materials in their three conditions were not identical at the sentence level, which suggests that the contrasts did not completely filter out some sentence-level processes. The biggest difference, however, is the absence of any context to accompany the question-answer pair. This, possibly, did not allow for expectations about the speaker's social goals, beliefs or motivations to build up, and therefore the use of contextual information by the participant was rather limited.

The study on indirectness by Shibata, Abe, Itoh, Shimada, and Umeda (2011) has in part the same conditions as the studies reported in Chapters 2 and 3, as they compared literal (direct) and face-saving indirect replies, plus an additional condition of contextually unconnected sentences. Just as in the design by Jang et al. (2013), there was a question (*"Can you work my shift this Friday?"*) and a reply (*"I am having a party on Friday"*), with two context sentences before the question. The questions as well as the replies in the different conditions were not matched in any way. In a comparison between the indirect and the direct ("literal") reply, there were activations in the bilateral medial frontal gyrus, bilateral inferior frontal gyrus (BA47 and BA9), and right middle temporal gyrus. Given the similarity of the design to the study in Chapter 2, one would expect overlap in the TPJ as well. One possible explanation why this region did not show up in the comparison is that the materials were not matched at the sentence level, and there was too much variability, adding noise and obscuring real effects. Although this is just a speculation, the right TPJ was activated in each of the three studies that I report here, so it is unlikely that it is just a spurious result, but is a fundamental part of the indirectness effect.

Lastly, Feng et al. (2017) turned to investigating the precise role of context in interpreting indirect utterances. They contrasted different ways of contextual revelance in a design with context stories and QA pairs. The target indirect utterances were either a) relevant, with replies obviously related to the context and the question, and all three repeating a key phrase for coherence; b) irrelevant with a contextual hint, or c) irrelevant without a contextual hint. The difference between the last two conditions was that even though the QA pair was the same (and less related than in the indirect relevant condition), there either was or was not a phrase linking the context to the reply, and thus making the interpretation easier. To give examples, a reply "*Nowadays, people are really beginning to enjoy opera*", could follow a question a) "*Do you think that more people are beginning to like opera?*" or b) *"Do you think that the audience liked my opera performance?"*. For the irrelevant condition without a contextual hint, the question was even less related (*"Do you think the audience will vote for my performance?"*), and the preceding story did not mention the word "opera", refering to an unspecified musical performance instead.

The main comparison of indirect to direct replies was largely overlapping with our activations. Interestingly, the effect of contextual relevance showed up in the right middle temporal gyrus, with an anterior-posterior gradient related to different levels of contextual relevance.

To conclude, it seems that there are indeed important variables influencing what regions are activated, even if the contrasts (indirect utterances > direct utterances) in these studies are superficially very similar. The most relevant variable seems to be whether there is richer local context against which the target utterance is interpreted, or not. With minimal contextual expectations, the activations are more similar to activations related to interpreting sentence-level meaning, with little or no right-hemisphere and/or mentalizing regions involvement. This suggests that sufficiently rich and natural experimental designs are crucial when investigating pragmatic phenomena – as I further elaborate on in the next section.

In addition, there is also related MEG and EEG work not strictly on indirectness, but on pragmatics and conversation, attesting to the general conclusion that mentalizing is recruited during utterance interpretation, and specifying the conditions for its recruitment. Bögels, Barr, Garrod and Kessler (2015) showed in an MEG study that mentalizing regions (TPJ and PFC) are not activated routinely, but when the need to update common ground arises, such as with pragmatic violations in the use of common ground.

What is also interesting in the context of speech act comprehension is the recent work on the timing of (direct) speech act recognition. In several experiments, Gisladottir (R. S. Gisladottir, 2015; Gísladóttir, Chwilla, Schriefers, & Levinson, 2012; Rosa S. Gisladottir, Bögels, & Levinson, 2018; Rosa S. Gisladottir, Chwilla, & Levinson, 2015; reviewed in Bögels & Levinson, 2017) used variations at the level of speech acts, e.g. comparing answers, pre-offers and declinations, while keeping the utterance identical. She showed that people recognize speech acts very early after the speaker's conversational turn starts, in some cases even before the actual words are uttered (see also Egorova, Shtyrov, & Pulvermüller, 2016).

4. Indirectness and experimental designs

As the studies in this thesis are taking place at the pragmatic level of analysis, it is important that the stimuli feel natural to the participants – otherwise, they might use strategies which deviate from those that they would engage in in real life (see Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012 for a similar argument and evidence about irony). I have tried to reach the best compromise between experimental control and naturalness of communicative situations in different ways in each study. In the following, I will describe the advantages and possible disadvantages of these designs. I will conclude by discussing to what extent naturalness of experimental designs and/or interactivity is necessary in studying indirect communication.

In Chapter 2, I used auditorily presented short context stories and four-turn dialogs, where the last reply was the target utterance. The advantage of this approach was that I could have substantial control over what is in the participants' common ground – i.e. the key information the participant was supposed to have for inferring the intended implicature from the target utterance. The downside of such approach is its predictability. Because the participants had all the key information "served" to them in the context story, they were having an easier task in making the implicature than in real-life conversations, where many pieces of information can be unknown, or unclear. In addition, even though I tried to achieve fairly natural intonation and timing of the entire dialog and the target utterances, it was not possible to include fillers (e.g. "*uhm*", "*well*") that typically accompany dispreferred conversational turns (e.g. Pomerantz, 1984), or to have the reply unfold over several conversational turns.

The setting created in the job interview study in Chapter 3 seems to be, on the surface, the most similar to a person's typical experience when having conversations – it was (seemingly) interactive, with multiple speakers, and with target utterances relevant to the participants, as they had to take their implicated meaning into account when making decisions about who is the best job candidate. On the other hand, believing that one is in a real-life, interactive, situation brings along additional assumptions. For example, there might be a strong expectation that one can ask for a clarification in case she did not fully understand a given reply. Because the questions had to be scripted for our paradigm to work, and the replies were recorded, it was not possible to provide such fully interactive experience. From this point of view, the paradigm was rather "minimally interactive" – carrying some key signs of interaction, such as turn-taking, but not fully reaching the complexities of real-life communicative exchanges.

In contrast to the job interview study, studies in Chapters 4 and 5, were kept deliberately simple in terms of the delivery of the indirect utterance. I used three-frame cartoons involving two protagonists, with the target sentence always on the final frame. Since the indirect messages in these two chapters were not primarily face-saving, but rather informative, the communicative exchanges were rather simple. There were also no issues on how to time the indirect answers relative to the questions, since the cartoons were presented via self-paced reading and participants timed the replies themselves via a button-press. In addition, effort was made to draw the facial expressions of both protagonists in a natural and believable way, since facial expressions also have a communicative function (e.g. Parkinson, 2005). The upside of using cartoons is that even though they are greatly simplified in comparison to real-life situations, they are an established 'overhearing' genre and therefore did not feel particularly unnatural to participants.

In general, even with effort taken to make the presentation of indirect replies fairly natural, their delivery was still different from that in real life communicative exchanges. When people talk to each other, there are a lot of things going on in every single moment: unfinished sentences and ungrammatical expressions, background noise, overlapping conversational turns, rapid changes in intonation and facial expressions, gesturing, and hidden impression management - to name just a few. Studying how people communicate entails controlling as many of these variables as possible, and it is inevitable that researchers make a number of simplifications, or "idealizations" (Clark, 1997) in the process. The most important general

adjustment I had to make in order to ensure sufficient experimental control was to limit the unpredictability that comes with truly natural conversations, and to resign on real interactivity. So, for example, my experimental paradigms did not reflect the cooperative nature of conversations: that speakers and recipients do not communicate in flawlessly crafted and perfectly understandable utterances, but co-construct the unfolding meaning together. Thus, an indirect reply that was always contained in a single target utterance in my studies would normally span several conversational turns, with the speaker and recipient mutually checking whether the message is clear and what are its implications (Clark, 1996).
At present, it is clear that with the technical limits imposed on experimental design and analysis with the current technology, we cannot simply put the participants in the fMRI scanner and let them freely speak to each other. The question to ask is – to what extent is it significant? Do the simplifications I have made in designing these studies invalidate the comparisons between direct and indirect communication in significant ways? From the processing point of view, in what way is inferring communicative intentions from language comprehension in real-life settings, such as in a dialog, different to a one-way monologue?

One feature that interactivity brings along is that typically, the listener has to be more engaged at many levels: not only decoding what was said, but also coordinating turn-taking and planning the next turn. At the same time, there are potentially a lot more sources of relevant context than in non-interactive monologues: for example, subtle extralinguistic cues such as gaze shifting, changes in body posture, facial expressions, etc. All these sources of information make the communicative situation richer, but also more complex. Without a clear experimental instruction, the listener is left to make her own calls as to which of these cues are relevant and which she should prioritize in case there is conflicting information (e.g. when somebody insults her with a seemingly friendly facial expression, as is the case with passive aggression). While this is potentially also the case for overhearing (e.g. watching a movie), being an *active* addressee uniquely adds to this complexity, since it requires extra activity, i.e. planning one's own utterance and continuing to show engagement in the interaction by back-channeling ("*uhm*", "*I see*", "*that was great!*").

Taken together, the task of the overhearer, and especially of the addressee, is much more complex and computationally challenging in real face-to-face conversations than in artificially constructed experimental designs. This does not automatically imply that the neural correlates of interpreting communicative intentions from indirect replies should be entirely (qualitatively) different if I studied them in real-life settings. Rather, it suggests that if there are any differences between being an addressee/overhearer of indirect remarks in real-life settings, as opposed to being a participant in the kind of experiments reported in this thesis, they are probably quantitative and not qualitative (one indication is that the indirectness effect did not substantially change from study to study, even though each contained a number of different simplifications). The most critical issue seems to be that in real life settings, there are many more sources of contextual information that must be taken into account, and this – together with a missing key as to which of these cues are relevant – makes interpretation more complex. In experimental settings, it is typically much clearer which aspects of the situation provide cues to interpreting the indirect remark.

5. Future directions

The above discussion suggests that one of the ways research in the neurobiology of language can be improved to better reflect the computations that are taking place during real-life communication is to make experimental paradigms as close as possible to real "arenas of language use" (Clark, 1992). Currently, research in pragmatics and experimental pragmatics is undergoing a "second revolution" (e.g. Bara, Enrici, & Adenzato, 2016), where the ambition is to move from strictly controlled "passive" experimental designs, through more naturalistic situations, all the way to direct engagement of the participant in a rich, multi-modal experience. This involves participants receiving momentary feedback and having the chance to co-construct the conversation they are part of. With the help of new technologies such as virtual reality, this is becoming increasingly possible. However, the current approaches are still rather general, and they do not enable researchers to control the important methodological aspects in at least a comparable manner as with non-interactive paradigms. The studies in this thesis are best positioned in the middle of the above continuum – the participants were fairly active and, in the job interview study, even directly engaged as addressees of indirect replies. While I gained important insights even without going "fully interactional", it is instructive to think about additional design aspects that would bring added value to the study of pragmatic inferences behind speaker meaning interpretation.

One way in which studying the pragmatic level of language use could be interestingly enriched is to enable participants to have something “at stake”. At present, most studies provide no real motivation for addressees to make their inferences as quickly and as efficiently as possible. While the active addressee condition in the job interview study did engage participants directly, they still had nothing to lose if they did not make their inferences quickly and accurately. Therefore, one way of engaging participants in a similar way as in real life would be to make their successes or failure in the task dependent on their speed and accuracy of drawing inferences and getting at the implicated meanings behind indirect utterances.

Another improvement, based on the discussion in the preceding section, would be to make the experimental situation less constrained, i.e. to add complexity in a systematic way. For example, to systematically vary different cues to interpretation – semantic relatedness of the question to the reply, prosody, facial expressions, status differences between conversational partners, the timing of gestures, etc. – and then observe which cues are contributing towards participants making their inferences more quickly and accurately.

Moving away from design issues to more conceptual questions, it would be also interesting to look at individual differences in indirect language interpretation, as well as their neural correlates and developmental trajectories. In the behavioral study in Chapter 4, there has been fairly large individual variation in the speed of indirect reply comprehension among participants; in addition, there are indications that both children (e.g. Wang, Lee, Sigman, & Dapretto, 2006) and adults (Holtgraves, 1997; Nieuwland, Ditman, & Kuperberg, 2010) substantially differ in how fast, how well, and how routinely, they draw on implicit meaning in speech (e.g. how fast they “get” irony or whether they have a tendency to look for hidden messages in other people’s utterances). Knowing what underlies this ability could potentially lead to devising strategies to help populations that have difficulties with implicit meaning interpretation, such as people on the autistic spectrum or dementia patients.

Another interesting issue which deserves more research is to directly compare pragmatic inferences behind speaker meaning interpretation with mentalistic inferences in ToM reasoning. In the field, this has not yet been done as ToM research is part of a different research tradition, focused on social cognition in general and not on language. The comparison between the indirectness effect and false belief localizer in Chapter 5 suggests that inferences behind false beliefs and indirect replies yield only partial overlap. A systematic comparison of different mentalizing tasks and various pragmatic phenomena (e.g. figurative language, indirect replies, irony) within the same group of participants would show to what extent are the inferences behind speaker meaning interpretation dependent on the same neural substrate as ToM reasoning, or which parts of the ToM network are implicated in which pragmatic phenomena.

Currently, there is almost no research on *gestural* indirectness apart from a small number of developmental studies reviewed in Chapters 4 and 5. In this thesis, I focused on the gap between getting at the gesture's referential and social intention. However, pointing gestures are also interesting in an "interactional" sense – when a speaker uses relatively infrequent means of communication, such as a pointing gesture instead of a verbal answer, she usually has a reason, e.g. wanting to express a playful attitude or letting the addressee know that the answer was obvious, if he only cared to search for it. One could contrasts the use of indirect gestures which do have such extra "interactional" aspect, with gestures motivated purely by the fact that the speaker's hands were occupied and therefore a gesture was the only means of communication. This would speak to the issue of whether, and how, social context modulates indirect meaning interpretation. Another way in which studying indirect pointing gestures can inform psycholinguistics is to zoom in on the role of facial expressions in interpreting the indirect message, and contrast it with the role of intonation in interpreting verbal utterances.

Lastly, while the fMRI technology is able to show which brain region is associated with which experimental contrast, it cannot prove causality. Therefore, the logical next step in investigating the neural correlates of indirect meaning interpretation would be to use technology that is suitable for this, for example TMS.

6. Conclusion

The studies presented in this thesis represent one of the first attempts in the field of the neurobiology of language to study particularized conversational implicatures, i.e. to focus at a purely pragmatic level of language use. At the same time, I treat language comprehension as driven by social motives and thus study communication as not only exchange of information, but also as means of navigating one's social world. In the following, I will recap the main insights that this work has delivered.

I have shown that - as predicted by theoretical accounts in pragmatics - mentalizing holds a significant role in the interpretation of speaker meaning, and that simple mirroring is not sufficient for this task. In directly contrasting the mentalizing network involved in interpreting false beliefs and indirect replies, I have also shown that these two are not entirely overlapping. This suggests that the pragmatic inferences involved in speaker meaning interpretation involve more "implicit" mentalizing than tasks where one has to explicitly think about the speaker's thoughts, desires or false beliefs. Contrasting indirect replies delivered via verbal utterances or pointing gestures revealed that the same mentalizing and discourse-level networks are involved in talking and pointing; however, interpreting indirect pointing gestures is still harder than their verbal counterparts because indirectness provides less constraints for the addressee to determine the focus of the gesture. I have also shown that interpreting indirect replies from the position of an overhearer is not fundamentally different from that of a direct addressee, especially when they have an identical goal in conversation. Finally, the difference between my results and other published work on indirectness suggests that in order to study speaker meaning, it is crucial to place utterances in sufficiently rich context.

References

Bara, B. G., Enrici, I., & Adenzato, M. (2016). Chapter 54 - At the Core of Pragmatics: The Neural Substrates of Communicative Intentions. In *Neurobiology of Language* (pp. 675–685). https://doi.org/http://dx.doi.org/10.1016/B978-0-12-407794-2.00054-7

Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In M. Beeman & C. Chiarrello (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience.* (pp. 255–284).

Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2015). Conversational Interaction in the Scanner: Mentalizing during Language Processing as Revealed by MEG. *Cerebral Cortex*, *25*(9), 3219–3234. https://doi.org/10.1093/cercor/bhu116

Bögels, S., & Levinson, S. C. (2017). The Brain Behind the Response: Insights Into Turn-taking in Conversation From Neuroimaging. *Research on Language and Social Interaction*, *50*(1), 71–89. https://doi.org/10.1080/08351813.2017.1262118

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language-A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, *50*(11), 2669–2683.

Clark, H. H. (1992). *Arenas of language use*. University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2016). Brain basis of communicative actions in language. *NeuroImage*, *125*, 857–867. https://doi.org/10.1016/j.neuroimage.2015.10.055

Feng, W., Wu, Y., Jan, C., Yu, H., Jiang, X., & Zhou, X. (2017). Effects of contextual relevance on pragmatic inference during conversation: An fMRI study. *Brain and Language*, *171*, 52–61. https://doi.org/10.1016/j.bandl.2017.04.005

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific Brain Networks during Explicit and Implicit Decoding of Emotional Prosody. *Cerebral Cortex*, *22*(5), 1107–1117. https://doi.org/10.1093/cercor/bhr184

Gisladottir, R. S. (2015). *Conversation electrified: The electrophysiology of spoken speech act* *recognition.* Radboud University, Nijmegen.

Gisladottir, R. S., Bögels, S., & Levinson, S. C. (2018). Oscillatory Brain Responses Reflect Anticipation during Comprehension of Speech Acts in Spoken Dialog. *Frontiers in Human Neuroscience*, *12*, 34. https://doi.org/10.3389/fnhum.2018.00034

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation Electrified: ERP Correlates of Speech Act Recognition in Underspecified Utterances. *PLOS ONE*, *10*(3), e0120068. https://doi.org/10.1371/journal.pone.0120068

Gísladóttir, R. S., Chwilla, D. J., Schriefers, H. J., & Levinson, S. . (2012). Speech act recognition in conversation: Experimental evidence. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*. Curran Associates, Inc.

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70–86. https://doi.org/10.1016/J.JML.2016.01.001

Holtgraves, T. (1997). Styles of language use: Individual and cultural variability in conversational indirectness. *Journal of Personality and Social Psychology*, *73*(3), 624–637. https://doi.org/10.1037/0022-3514.73.3.624

Holtgraves, T., & McNamara, P. (2010a). Parkinson's Disease and Politeness. *Journal of Language and Social Psychology*, *29*(2), 178–193. https://doi.org/10.1177/0261927X09359521

Holtgraves, T., & McNamara, P. (2010b). Pragmatic comprehension deficit in Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *32*(4), 388–397. https://doi.org/10.1080/13803390903130729

Jang, G., Yoon, S., Lee, S.-E., Park, H., Kim, J., Ko, J. H., & Park, H.-J. (2013). Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. *NeuroImage*, *81*, 61–72. https://doi.org/10.1016/j.neuroimage.2013.05.027

Kissine, M. (2015). Pragmatics as Metacognitive Control. *Frontiers in Psychology*, *6*, 2057.

https://doi.org/10.3389/fpsyg.2015.02057

McNamara, P., Holtgraves, T., Durso, R., & Harris, E. (2010). Social cognition of indirect speech: Evidence from Parkinson's disease. *Journal of Neurolinguistics*, *23*(2), 162–171. https://doi.org/10.1016/J.JNEUROLING.2009.12.003

Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (2009). When Elephants Fly: Differential Sensitivity of Right and Left Inferior Frontal Gyri to Discourse and World Knowledge. *Journal of Cognitive Neuroscience*, *21*(12), 2358–2368. https://doi.org/10.1162/jocn.2008.21163

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*(3), 324–346. https://doi.org/10.1016/J.JML.2010.06.005

Parkinson, B. (2005). Do Facial Movements Express Emotions or Communicate Motives? *Personality and Social Psychology Review*, *9*(4), 278–311. https://doi.org/10.1207/s15327957pspr0904_1

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shaped. *Communication Faculty Scholarship*.

Rapp, A. M., Mutschler, D. E., & Erb, M. (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage*, *63*(1), 600–610. https://doi.org/10.1016/j.neuroimage.2012.06.022

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/J.TICS.2005.11.009

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Shibata, M., Abe, J. ichi, Itoh, H., Shimada, K., & Umeda, S. (2011). Neural processing associated with comprehension of an indirect reply during a scenario reading task. *Neuropsychologia*, *49*(13), 3542–3550. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2011.09.006

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39. https://doi.org/https://doi.org/10.1016/j.neuroimage.2012.06.046

St George, M., Kutas, M., Martinez, A., & Sereno, M. I. (1999). Semantic integration in reading: Engagement of the right hemisphere during discourse processing. *Brain*, *122*(7), 1317–1325.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Tylén, K., Wallentin, M., & Roepstorff, A. (2009). Say it with flowers! An fMRI study of object mediated communication. *Brain and Language*, *108*(3), 159–166. https://doi.org/10.1016/j.bandl.2008.07.002

Van Berkum, J. J. A. (2015). Language comprehension and emotion: where are the interfaces, and who cares? In M. Miozzo, G. de Zubicaray, & N. O. Schiller (Eds.), *Oxford Handbook of Neurolinguistics*. Oxford: OUP.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain*, *129*(4), 932–943. https://doi.org/https://doi.org/10.1093/brain/awl032

SAMENVATTING

Nederlandse samenvatting

Elkaar verstaan lijkt een eenvoudig en rechttoe-rechtaanproces. We zitten met een idee in ons hoofd, verpakken het in woorden, lijmen ze samen met behulp van wat grammatica en sturen dit pakketje richting de toehoorder. Hij of zij decodeert dat pakketje, zoekt de woorden op in het langetermijngeheugen, gebruikt aangeleerde grammaticaregels om een coherente boodschap te formuleren en begrijpt perfect wat we willen zeggen. Maar dit simpele beeld wordt snel ingewikkeld als we de volgende gevallen onder de loep nemen:

Als een ziekenhuisverpleegster een college informeert dat “de appendix in kamer 4 nog haar avondmaal moet nemen”, dan is er een wederzijds begrip dat niet precies de blinde darm honger heeft, maar uiteraard een patiënt die met appendicitis gehospitaliseerd ligt. Wij zeggen “ik heb het koud” op een date, wat een neutrale mededeling kan betekenen maar ook een uitnodiging om te knuffelen kan zijn. Een wijzend gebaar naar het lege voederbakje van de kat kan bedoeld zijn als herinnering voor een kamergenoot die problemen heeft om de regelmatige eettijden van de kat in de gaten te houden. En als je hoort “het is toch moeilijk om een goede presentatie te geven” van een collega die het goed met jou meent, kan dat simpelweg een statement zijn over de moeilijkheden van Powerpoint te gebruiken, maar ook een feedback met nogal wat medelijden na jouw allesbehalve ideale optreden op een conferentie. We *zeggen* dus duidelijk iets anders dan we *willen* communiceren.

In deze thesis heb ik gekeken hoe ons brein erin slaagt om uitingen te decoderen, waarin de geuite zin een interpretatie bevat die verschillend is en in sommige gevallen zelfs het tegenovergestelde tot wat de spreker echt gecommuniceerd wil krijgen. Als een reeks voorbeelden om op te focussen, koos ik voor indirecte reacties, zoals het voorbeeld van de niet-zo-geweldige presentatie die eerder vermeld werd.

Waarom spreken mensen op indirecte wijze? Hier zijn verschillende redenen voor, maar, zoals mijn voorbeeld impliceert, vermijden we dikwijls dingen direct te zeggen omdat we beleefd willen zijn voor de persoon met wie we interageren. Mocht er iemand uitflappen “Je presentatie was echt een zootje!”, dan zou hij zeker zijn punt gemaakt hebben, maar ten koste van de gevoelens van zijn of haar vriend en wellicht met reputatieschade als hij of zij bekend staat als een meelevend iemand. Gevoelige zaken op indirecte manier communiceren zorgt ervoor dat de interpretatielast overgebracht wordt naar de luisteraar en zorgt er dus voor dat de spreker geen “gezichtsverlies” lijdt. Hij houdt er dus een positief zelfbeeld aan over.

Hoe heb ik in het brein gekeken terwijl we de echte betekenis van de indirecte reacties probeerden uit te vissen? Ik heb hersenbeeldtechnologie gebruikt die we functionele MRI of functionele kernspintomografie noemen (fMRI). fMRI is een machine met een opening zoals een tunnel, waarin de patiënten liggen terwijl ze een taak uitvoeren – zoals naar zorgvuldig opgebouwde zinnetjes luisteren – en met behulp van krachtige analysetools kan gekeken worden naar wat er zich in hun hersenen afspeelt. Deze “inkijk” in hun hersenactiviteit is gebaseerd op het feit dat wanneer een bepaald gedeelte van de hersenen betrokken wordt bij een taak, dit deel energie nodig heeft om de taak uit te voeren. fMRI slaagt erin om veranderingen in het bloed, dat de energie levert aan dat gedeelte van de hersenen waar de taak uitgevoerd wordt, op te merken.

Hoe verwerken onze hersenen dus indirecte reacties? Ik zal deze vraag beantwoorden door te verwijzen naar vier experimenten, een voor een uitgevoerd in onze dissertatie.

Aangezien er geen gelijkaardig onderzoek ter beschikking stond ten tijde van het begin van mijn onderzoek, wilde ik aanvankelijk eenvoudigweg een brede context scheppen en een basisvergelijking maken tussen reacties die hetzelfde klinken, maar ofwel eerder een directe lezing (bijvoorbeeld na de vraag: “Hoe moeilijk is het om een goede presentatie te geven?) of een indirecte betekenis (“vond je mijn presentatie overtuigend?”) bevatten. Om nog verder te reflecteren over hoe mensen indirectheid in hun alledaags taalgebruik inbouwen, zorgde ik ook voor indirecte reacties zonder een motief van mogelijk gezichtsverlies. Gevallen waar de spreker enkel het antwoord meer informatief wilde maken en daardoor reageerde op een ja/neenvraag door de mogelijke redenen voor de “ja” of “neen” al op voorhand te geven (“Zal je een poster of een presentatie voor je conferentie kiezen? Het is moeilijk om een goede presentatie te geven”). Door zinnen te vergelijken die hetzelfde klinken maar iets anders betekenen dankzij de context waarin ze gebruikt worden, kon ik bestuderen welke gebieden in de hersenen (meer) betrokken zijn bij het verwerken van elk type reactie. Ik heb vastgesteld dat naast de gebieden in de linkerhersenhelft, die betrokken zijn bij het decoderen van woordbetekenissen en grammaticale relaties onderling, er ook activiteit te zien was in gebieden die betrokken zijn bij contextuele integratie. Dat betekent dat als we willen begrijpen wat de indirecte reactie precies betekent, de luisteraar niet enkel de woordbetekenissen van de reactie die opgeslagen zijn in het langetermijngeheugen moet activeren, maar ook zijn of haar kennis van de wereld over presentaties, academisch succes, eindexamens en enkele verzwegen communicatieafspraken moet gebruiken. Cruciaal hierbij was dat ik ook gebieden geregistreerd heb die het perspectief van de ander kunnen begrijpen en impliceren. Dit wordt de "theory of mind" genoemd, het vermogen zich te verplaatsen in de gedachten van anderen. Dat wil zeggen als we willen begrijpen wat mensen echt bedoelen als ze indirect spreken , we ook hun overtuigingen over de situatie moeten beschouwen – bijvoorbeeld dat ze op een bepaalde manier zouden kunnen spreken omdat dit een gevoelig thema is en dat ze onze gevoelens niet willen kwetsen en tegelijk ons nog feedback willen geven.

In mijn tweede experiment kwam ik dichter bij reëel taalgebruik door een situatie te creëren waar een indirect antwoord gericht zou zijn *naar* de deelnemer *toe*. De gedachte hierachter lag in het idee dat in de eerste gevalstudie de deelnemers zich gedroegen als toehoorders, die naar verhalen en dialogen over andere mensen en hun levens luisterden. Dit is niet hoe we gebruikelijk functioneren en ik was benieuwd of er enkele veranderingen zouden zijn in de hersenactiviteit, mocht de deelnemer direct betrokken zijn als ontvanger van indirecte antwoorden. Om een realistische setup te creëren waar het mogelijk is om veel indirecte reacties te krijgen (dit is een noodzakelijke voorwaarde om een goed fMRI-signaal te krijgen), ontwierp ik een fictief jobinterview. Een jobinterview is een goede omgeving waar mensen op natuurlijke wijze indirect taalgebruik vertonen om hun troeven uit te spelen en af en toe zichzelf wat beter afschilderen dan ze zijn, of ze proberen een tekort aan relevante eigenschappen en ervaring te maskeren. In de studie werden de deelnemers in de fMRI-scanners gebruikt als interviewers die de vragen moesten stellen en antwoorden kregen van verschillende kandidaten voor de job, die ze zouden moeten evalueren aan het eind van dit interview. Sommige antwoorden waren direct ("wat zijn je plannen voor de zomer?" "Ik ben van plan deze zomer deel te nemen aan een zomercursus.") en sommige indirect ("spreekt u vloeiend Engels?" "Ik ben deze zomer van plan deel te nemen aan een zomercursus"). Aangezien we een situatie wilden die vergelijkbaar zou zijn met studie 1, werd er ook een controlevoorwaarde ingebouwd waarbij deelnemers luisterden naar indirecte en directe uitingen als toehoorders, die zouden opgenomen zijn op de dag voor de interviews van dezelfde kandidaten met een vorige deelnemer aan de studie.

Omdat ik een zuivere vergelijking wilde tussen hoe actieve deelnemers en hoe toehoorders indirecte reacties verwerken, moest ik zeker zijn dat alle factoren van de reacties dezelfde waren en er enkel verschillen zaten in de actieve/passieve rol van de deelnemer. Daarom stonden alle vragen die de deelnemers moesten op script en had ik de antwoorden van de kandidaten (onderzoeksassistenten) vooraf opgenomen, om identieke timing en intonatie voor alle deelnemers te verzekeren. In de resultaten van de slotanalyse voegde ik enkel de deelnemers in die geloofden dat ze in real time met de kandidaten aan het praten waren.

Het resultaat bevestigde opnieuw dat nog bijkomend bij het kerntaalnetwerk dat gelocaliseerd moet worden in de linkerhersenhelft, indirecte reacties geactiveerd worden in gebieden in de rechterhersenhelft, die in verband staan met het nemen van perspectief en contextuele integratie – net als in de eerste gevalstudie. Wat was er anders wanneer deelnemers *directe ontvangers* waren van indirecte reacties om gezichtsverlies te vermijden? In tegenstelling tot wat we verwacht hadden, waren hier geen gebieden actief die in verband staan met de “theory of mind” of contextuele integratie, maar met emotionele karaktertekening – wellicht niet omdat het moeilijker wordt om de bedoeling van de spreker te achterhalen als iemand direct betrokken is in een jobinterview, maar omdat men meer emotionele betrokkenheid voelt als de spreker probeert een indruk te maken op *mij*, en *mijn* mening over hem of haar probeert te beïnvloeden.

In de laatste twee gevalstudies bleef ik werken met indirecte reacties maar ik heb de modaliteit gewijzigd waarin ze uitgesproken werden. In *real-life* communicatie, maken mensen dikwijls gebruik van allerlei soorten communicatieve signalen, niet enkel van woorden. Met voldoende context kan om het even wat gebruikt worden om te communiceren – een knikje, een glimlach, een goed getimede stilte of zelfs een simpel wijzend gebaar. Daarom maakte ik in de laatste twee projecten *cartoon* scenario´s waar een persoon een vraag zou stellen en de andere persoon een direct of indirect antwoord zou leveren met ofwel een zin of een handgebaar. Bijvoorbeeld een hardloper die zijn huisgenoot zou vragen of hij zou willen meedoen en de huisgenoot zou ofwel zeggen: “in mijn hardloopschoen zit een gaatje,” of – in de wijsvoorwaarde – enkel het wijzen met zijn vinger naar de beschadigde hardloopschoen. Ik heb vergeleken hoe snel de deelnemers konden vinden wat het gebaar betekende in vergelijking met een zin met identieke betekenis (gevalstudie 3) en ook welke hersengebieden er betrokken zijn bij zowel gebaar- als woordelijke indirectheid, en welke hersengebieden verschillen (studie 4) .

In de derde studie concludeerde ik dat de deelnemers even snel zijn om de directe reacties te interpreteren – het maakte niet uit of ze met een handgebaar of een zin geuit werden. Er was echter een grote vertraging om indirecte boodschappen in gebaren te decoderen in vergelijking met verbale indirectheid. Mijn interpretatie van dit effect is dat indirecte gebaren altijd meer ambigu zijn dan indirect spraakgebruik, omdat de luisteraar moeilijkheden kent om niet enkel de boodschap te ontcijferen, maar ook welk deel van het signaal uiteindelijk relevant is – is het de volledige schoen, is het het specifieke merk van de schoen, het gaatje in de schoen of zelfs nog iets anders. Niet-verbale communicatieve signalen zijn krachtig, maar door verbale expresssie te gebruiken maken we het overbrengen van complexe boodschappen gemakkelijker.

Deze veronderstelling werd bevestigd in de fMRI-versie van deze taak, waar ik nogmaals het algemeen patroon van de activering van indirectheid kon zien uit studies 1 en 2, en een specifiek verschil voor gebaren en indirectheid in het gebied van de hersenen dat betrokken is bij het verwerken van niet-conventionele signalen met een ambigue betekenis.

Over het algemeen tonen de studies die ik uitgevoerd heb in mijn dissertatie aan dat luisteraars heel flexibel zijn in het gebruiken van contextuele informatie, inclusief het perspectief van de spreker, om te belanden bij de bedoelde boodschap van de spreker – zelfs als het de tegenovergestelde betekenis heeft van wat er "in" de woorden zelf verborgen ligt. Wat ook duidelijk wordt, is dat het kerntaalnetwerk dat zich in de linkerhersenhelft van de hersenen bevindt niet "het hele verhaal" kan verklaren – om te komen tot indirecte betekenissen van boodschappen, moeten we context raadplegen waar ook communicatie in vervlochten is, inclusief onze kennis over de spreker en zijn of haar intenties. Naast inzicht geven in een basisonderzoek neurobiologie bij taalgebruik, kan dit werk potentieel informatief zijn om mensen te helpen met communicatiestoornissen, vooral bij pragmatieke taalstoornissen, zoals autismespectrumstoornis en frontotemporale dementie.

Acknowledgments

Finishing this thesis took me much longer than planned. A lot happened during that time, and most of it was located more than a thousand kilometers from Nijmegen. Hanka and Samko were born and raised in a culture which does not believe in day-care for children under three. I held a full-time scientific job in experimental psychology at my Slovak institute. I spent countless hours teaching undergraduate students. I collected the data for three out of the four studies reported here during week- and weekend-long trips while there was always at least one attachment-parented toddler waiting for me at home. Despite all this, I am very happy I did not give up. Right now, seeing the final draft of my thesis on the computer screen, feels equally great and surreal.

All this would certainly not have been possible without the help and good will of a lot of people around me. I want to thank all of you, because this thesis is partly your achievement as well.

First of all, I want to thank Peter Hagoort and Jos van Berkum. I simply cannot imagine more supportive, patient, encouraging and just overall nice supervisors. Peter, you believed in me despite the slow progress and delays, and you always found ways to make the whole journey more manageable. I cannot thank you enough for this and when I have my own PhD students, the bar is set very high for me.
Jos, there were so many times when you said exactly what I needed to hear to not give up. I loved brainstorming about new ideas with you, admired your creative input and appreciated the ability to put things into perspective when I was overwhelmed. I knew that I could always count on your empathy and kindness, even if my issues were not strictly work-related. Thank you!

Kirsten and Katinka, my paranymphs. You are amazing. I feel that I cannot even specify the myriad of ways that you have supported me over the years. You shared your houses with me, cooked me dinners and made me countless cups of tea when I was down. You laughed and cried with me. You are great friends and I hope that I will be able to “return” your kindness one day.

My fellow former CNS Master classmates – Jonas, Gio, Marieke, Michael, Annemarie, Iris…. I liked our two years in Nijmegen. You haven't changed a bit ☺. Thank you for keeping in touch and being supportive!

Thanks to everybody who provided me with their couches and rooms to crash on while I was in Nijmegen – Kirsten, Katinka, Miriam de Boer, Saskia, Michael Klein… My couch is always there for you ☺, just come and visit.

I am also thankful to the people who helped me with the stimuli for some of my (pretty complicated) study designs. For Study 1, my thanks goes to pretty much all the student assistants, research assistants, technical staff, PhDs, postdocs, and even some of the PIs who helped me record the indirect dialogs. Thanks for spending your time in the miniature recording booth on the second floor just out of your good will! The same goes for the "narrator", Liesbeth Jansen, who sat through some of her Saturdays in the very same booth, patiently re-recording the background stories for my dialogs. Thanks to the MPI research assistants who recorded and then re-recorded the dialogs for the Job interview - Laura Arendsen, Brenda Lelie, José Kivits and Marjolijn van Gelder. Thanks to Katka Slaninková for drawing the cartoons for Studies 3 and 4. And Mirko Kolárik, thanks for the photo session with Lego and a plastic brain at -4°C at the main square in Bratislava.

Paul Gaalman – thanks for being such a fun (and kind) MRI lab manager. Scanning was never boring when you were around! Also, thanks to all my scanning buddies who were willing to sit with me in the DCCN basement during weekends and evenings. Especially since I know that the very idea of a scanning buddy is partly my fault ☺ (at least that's what Paul made me believe).

Thanks to Tildie Stijns, Ina Grevel and Carolin Lorenz – you also took great care of me and were always helpful.

I want to thank Prof. Viera Bačová and Peter Halama, who were the directors at the Institute of Experimental Psychology where I worked while also carrying out the research reported here. Thank you for always showing me appreciation and trust. Also, thank you for making the Institute such a welcoming, nice place for everybody.

Thank you, my Slovak colleagues and friends. Vlaďa and Lenka, you are such inspiration! Katka, Ivan, Marek, Lucia, Dáša, Lucia O., Jakub and everybody else from the fifth and fourth floor – work is never boring when you are around. Thanks to Janka Zemandl and Evka Vavráková, my friends and co-authors. You were always there for me.

Thanks to my running buddies from AKV. You probably have no idea what I was working on since we mostly just talk about running shoes, intervals and heart rates ☺, but our exhausting practices and races helped me keep my sanity.

My many students of Psycholinguistics – I know that I was the teacher, but you taught me a lot as well.

It takes a village to raise a child, and I want to thank my “village” for helping me with kids when I needed to spend the weekends working or travelling. My mom, Maja, babka Gabika and dedko Ivan – I am happy my kids have you as grandparents.

I also want to thank my Dad, who died when I was a teenager, but who managed to teach me – in the short time we had together – that knowledge matters. I think about you every day and would love to discuss my research with you.

Finally: Hanka and Samko – thanks for being such fun and happy kids. You probably don’t remember, but I took you to your first conferences abroad when you were 4 and 3 months old. And you managed pretty well ☺. Hopefully that inspired the future scientists in you! (And thanks for the Lego figures for the cover page!)

Ivan, I know that this has been hard for you as well. We met just before I started the CNS Master and then the PhD, and you have never seen me *not* being a workaholic. Thank you for your love and support. I promise to limit myself to one job at a time from now on ;).

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda Van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*
17. Moving eyes and naming objects. *Femke van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*

29. At the same time…: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Broersma*

35. Retrieving semantic and syntactic word properties. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*

43. Affix reduction in spoken Dutch. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*

45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*

50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*

51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*

52. Prosodic structure in speech production and perception. *Claudia Kuzla*

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*

54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*

55. Processing the fine temporal structure of spoken words. *Eva Reinisch*

56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*

57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*

58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*

59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*

60. Processing casual speech in native and non-native language. *Annelie Tuinman*

61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Robinson*

62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*

63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*

64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*

65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*

66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*

67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*

68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*

69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*

70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*

71. Gaze behavior in face-to-face interaction. *Federico Rossano*

72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Conny de Vos*

73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*

75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*

76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*

77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*

78. Contributions of executive control to individual differences in word production. *Zeshu Shao*

79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*

80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*

81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*

82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*

83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*

84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*

85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*

86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*

87. The acquisition of morphophonological alternations across languages. *Helen Buckler*

88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*

89. Rediscovering a forgotten language. *Jiyoun Choi*

90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*

91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*

92. Information structure in Avatime. *Saskia van Putten*

93. Switch reference in Whitesands. *Jeremy Hammond*

94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*

95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. *Beyza Sümer*

96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*

97. Incrementality and Flexibility in Sentence Production. *Maartje van de Velde*

98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*

99. The request system in Italian interaction. *Giovanni Rossi*

100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*

101. Assessing birth language memory in young adoptees. *Wencui Zhou*

102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*

103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*

104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition. *Rósa Signý Gisladottir*

105. Modelling Multimodal Language Processing. *Alastair Smith*

106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*

107. Situational variation in non-native communication. *Huib Kouwenhoven*

108. Sustained attention in language production. *Suzanne Jongman*

109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntatic, and orthographic effects. *Malte Viebahn*

110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*

111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*

112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*

113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*

114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*

115. Music and language comprehension in the brain. *Richard Kunert*

116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*

117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*

118. Language processing in a conversation context. *Lotte Schoot*

119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*

120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*

121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*

122. Sensorimotor experience in speech perception. *Will Schuerman*

123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*

124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*

125. Lapse organization in interaction. *Elliott Hoey*

126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*

127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*

128. The role of voice and word order in incremental sentence processing. *Sebastian Sauppe*

129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*

130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*

131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*

132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*

133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*

134. Indirect request comprehension in different contexts. *Johanne Tromp*

135. Envisioning Language - An Exploration of Perceptual Processes in Language Comprehension. *Markus Ostarek*

136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*

137. Let the agents do the talking: on the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*

138. Age and hearing loss effects on speech processing. *Xaver Koch*

139. Vocabulary knowledge and learning: Individual differences in adult native speakers. *Nina Mainz*

140. The face in face-to-face communication: Signals of understanding and non-understanding. *Paul Hömke*

141. Person reference and interaction in Umpila/Kuuku Ya'u narrative. *Clair Hill*

142. Beyond the language given: The neurobiological infrastructure for pragmatic inferencing. *Jana Bašnáková*

LOREM IPSUM

Lorem ipsum