

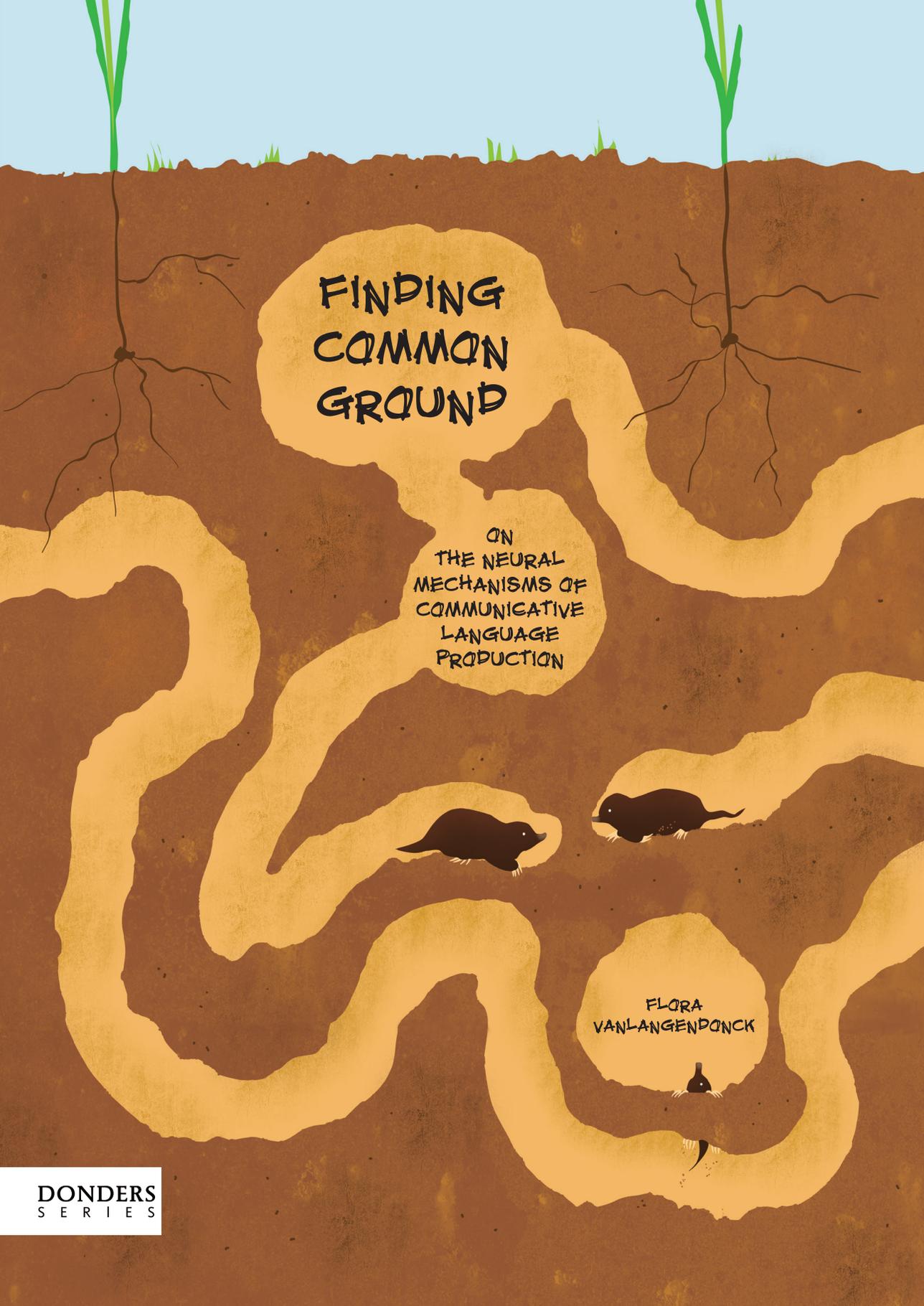
## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/167549>

Please be advised that this information was generated on 2019-08-21 and may be subject to change.



FINDING  
COMMON  
GROUND

ON  
THE NEURAL  
MECHANISMS OF  
COMMUNICATIVE  
LANGUAGE  
PRODUCTION

FLORA  
VANLANGENDONCK

DONDERS  
SERIES

Finding common ground

On the neural mechanisms of communicative  
language production

Flora Vanlangendonck

Finding common ground: On the neural mechanisms of communicative language production

PhD thesis, Radboud University Nijmegen

© Flora Vanlangendonck, 2017

ISBN 78-94-6284-101-7

Layout & cover artwork by Remco Wetzels - [www.remcowetzels.nl](http://www.remcowetzels.nl)

Printed by Ridderprint - [www.ridderprint.nl](http://www.ridderprint.nl)

This research was funded by a TopTalent grant from the Donders Graduate School for Cognitive Neuroscience.

Finding common ground  
On the neural mechanisms of communicative  
language production

**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 29 maart 2017  
om 12.30 uur precies

door

**Flora Vanlangendonck**

geboren op 17 augustus 1987  
te Jette (België)

**Promotor:** Prof. dr. Peter Hagoort

**Copromotor:** Dr. Roel Willems

**Manuscriptcommissie:**

Prof. dr. Ivan Toni

Prof. dr. Susan Brennan (Stony Brook University, Verenigde Staten)

Prof. dr. Simon Garrod (University of Glasgow, Verenigd Koninkrijk)

## Table of contents

<b>Chapter 1</b>	Introduction	9
<b>Chapter 2</b>	An early influence of common ground during speech planning	21
<b>Chapter 3</b>	Taking common ground into account: Specifying the role of the mentalizing network in communicative language production	45
<b>Chapter 4</b>	Collaborative learning engages different memory retrieval networks	71
<b>Chapter 5</b>	The role of the mentalizing network in the establishment of conceptual pacts	91
<b>Chapter 6</b>	General discussion	117
<b>Appendix</b>	References	129
	Nederlandse samenvatting	137
	Curriculum Vitae	144
	Publication list	145
	Acknowledgements	147
	Donders series	151



To my parents, Valentine and Geert



Introduction





## Introduction

Think of the last time you said something. In all likelihood you were talking to *someone* rather than muttering words to yourself. Whether you are gossiping, ordering a beer at a bar or teaching a class, the act of speaking typically involves a speaker and at least one addressee. It is therefore unsurprising that the communicative context in which we speak affects the way we speak. Our language system is very flexible in this respect, allowing speakers to adapt their language use depending *who* they are talking to, *why* they are talking and *what* they are talking about. For instance, if we discuss the topic of this thesis, I will adjust what I tell you and how I formulate it depending on what I know about you. If you do not speak Dutch, I will address you in English. If you are an expert in the field, I can use jargon. If we have talked about this research before, I may refer to our previous interactions. Thus, seemingly effortlessly, we can shape the language we speak to serve our and other people's communicative needs and goals.

This thesis consists of a series of experiments that investigate the cognitive and neural mechanisms that allow us to adapt our language use to the communicative context in which we find ourselves. In these experiments, I compare language production in communicative and non-communicative situations, as well as in situations in which speakers need to adjust their language use based on what information they do or do not share with their interlocutor.

### Language and communication

Given that we primarily use language to communicate, it is unsurprising that both language production and comprehension are shaped by communicative and social factors. A clear example of this is *recipient design* or *audience design* (Clark & Murphy, 1982; Sacks, Schegloff, & Jefferson, 1974), which refers to speakers' ability to tailor utterances so particular addressees can understand them. For example, you would probably not speak in the same way to a child as compared to an adult. Research has shown that speakers adapt their language use based on what they know about their addressee in a variety of ways (see e.g., Brennan, Galati, & Kuhlen, 2010 for an overview). For instance, speakers include less detail, use fewer words and produce less intelligible speech when retelling a story to an old compared to a new addressee (Galati & Brennan, 2010), and even four-year-old children produce shorter and simpler utterances when talking to a younger child compared to an adult (Shatz & Gelman, 1973).

In order to successfully tailor their language use to their audience, speakers need to take into account *common ground*, i.e. the beliefs, assumptions and knowledge they share with their interlocutors (Clark & Carlson, 1982; Clark & Marshall, 1981). Information that is only available to the speaker is referred to as *privileged ground* information. There has been considerable

debate in the literature about the extent and the way in which speakers take the distinction between common ground and privileged ground into account when designing utterances. Some theories assume that common ground is essential for efficient communication and that interlocutors continuously update their common ground (Clark & Brennan, 1991; Clark & Carlson, 1982; Clark & Krych, 2004). According to Constraint-Based Processing models, speakers keep their addressee in mind from the earliest stages of utterance planning in a probabilistic, constraint-based way, resulting in early effects of common ground (Brennan & Hanna, 2009; Hanna, Tanenhaus & Trueswell, 2003; Horton & Gerrig, 2002; Tanenhaus & Trueswell, 1995). In contrast, the Monitoring and Adjustment theory proposes that keeping track of the distinction between shared and privileged information during utterance planning is resource-intensive and often not necessary (Horton & Keysar, 1996). According to this view, speakers initially design their utterances from their own egocentric perspective, and common ground only comes into play at a later stage when speakers monitor their utterance (Keysar, Barr & Horton, 1998). In chapters two and three of this thesis, I investigate how and when common and privileged ground information affect speech planning.

Conversational partners keep track of moment-to-moment changes in common ground through a process known as *grounding*, i.e. they seek and give evidence that what has been said has been understood and thus make it part of their common ground (Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). For example, when a speaker introduces a new referring expression into a conversation (e.g., *did you see that car?*), the addressee can respond by accepting the expression (“yes”), modifying it (“*the red car?*”), or asking for a different conceptualization (“*what car?*”; Clark & Brennan, 1991). Once interlocutors have found and agreed on a shared perspective, they have established a *conceptual pact*, i.e. a flexible, temporary agreement about how a referent is conceptualized (Brennan & Clark, 1996). Interlocutors are likely to reuse conceptual pacts in later references.

There is evidence that conceptual pacts are partner-specific. For example, speakers reconceptualize conceptual pacts when switching conversational partners (e.g., Brennan & Clark, 1996; Horton & Gerrig, 2005), and listeners are slowed down in their interpretation when speakers break a previously established conceptual pact (Metzing & Brennan, 2003). However, the cognitive and neural mechanisms that bring about such partner-specific effects are still poorly understood. Clark & Marshall (1978; 1981) proposed that interlocutors encode triple co-presence of speaker, addressee and referent in specialized *reference diaries*. However, Horton & Gerrig (2005, 2016) pointed out that routinely making inferences about triple co-presence is too computationally costly to be plausible. Instead, they proposed that domain-general episodic memory processes support the emergence of partner-specific effects (Horton & Gerrig, 2005; Horton, 2007). In this view, conversational partners create rich episodic memory traces of their interactions, thus linking interlocutors, content and context

in memory. The presence of an interlocutor can then lead to the automatic retrieval of related information, including previously established conceptual pacts. This association-based view requires neither explicit tagging of common ground, nor specialized memory processes (Horton & Brennan, 2016). In the fourth and fifth chapters of this thesis, I investigate how conceptual pacts are established and retrieved from memory.

## **The neurobiology of communication**

Although we usually speak in social and communicative contexts, the neurobiology of speaking, and language processing in general, has typically been studied in isolation. For example, participants in these experiments may be instructed to read or listen to words or sentences, or to name pictures in settings stripped of any communicative or social context. One obvious reason for this is that it is hard to study social and communicative language use in the lab (see the notes on methods below). Yet, by ignoring the communicative and social context in which we use language, we risk missing out on key aspects of everyday, naturalistic language processing. In recent years, researchers have therefore started to investigate the neural mechanisms supporting the production and comprehension of communicative actions and language.

Neuroimaging studies on communication have studied a wide variety of verbal and non-verbal behaviors, including pointing, iconic gestures, language and the development of novel communication systems. Research on communicative action planning has consistently found that the right superior temporal sulcus is more activated when planning a communicative as compared to a non-communicative action. For example, the superior temporal sulcus is involved when planning a novel communicative action as compared to an instrumental action (Noordzij et al., 2009), and during communicative pointing relative to non-communicative pointing (Cleret de Langavant et al., 2011). Furthermore, repetitive TMS to this brain region impairs communicators' ability to improve their task efficiency over the course of a series of communicative interactions (Stolk et al., 2014).

The superior temporal sulcus also plays an important role in the comprehension of communicative actions. For example, Noordzij et al. (2009) showed that the right posterior superior temporal sulcus is involved in both the planning and the recognition of communicative actions, and the bilateral superior temporal sulcus is more activated while observing communicative, participant-directed gestures relative to non-communicative actions (Redcay, Velnoskey, & Rowe, 2016). Similarly, seeing someone shift their eye gaze towards you, a powerful way to initiate communication, has also been found to selectively engage the right posterior superior temporal sulcus (Ethofer, Gschwind, & Vuilleumier 2011; Redcay et al., 2010; Redcay, Ludlum, Velnoskey, & Simren, 2015). However, a number of studies have also found activations in a more distributed network of brain regions,

including key areas involved in mentalizing and theory of mind (see box 1). For example, the comparison between observing actions with a communicative intent as compared to a private intent revealed activations in the medial prefrontal cortex (Walter et al., 2004; Ciaramidaro et al., 2007; Ciaramidaro, Becchio, Colle, Bara, & Walter, 2014), temporoparietal junction (Ciaramidaro et al., 2007; Ciaramidaro et al., 2014) and the precuneus (Ciaramidaro et al., 2007). Similarly, establishing joint attention through eye gaze has also been found to activate the medial prefrontal cortex (Kampe, Frith, & Frith, 2003; Schilbach et al., 2010), the temporoparietal junction (Caruana, Brock, & Woolgar, 2015; Redcay et al., 2010) and the precuneus (Caruana et al., 2015).

A similar picture emerges from the literature on communicative language processing. Relatively few neuroimaging studies have manipulated communicative intent in a language production paradigm. Sassa et al. (2007) found that the medial prefrontal cortex, bilateral superior temporal sulcus and the left temporoparietal junction were more activated while speakers were talking to an actor relative to describing a situation. Willems et al. (2010) independently manipulated communicative intent and linguistic difficulty in a language production task. They found that the medial prefrontal cortex is sensitive to communicative intent, regardless of linguistic difficulty. The medial prefrontal cortex also appears to play an important role in the comprehension of communicative speech. For instance, the medial prefrontal cortex and bilateral temporal poles are more activated when someone hears their own name as compared to someone else's name (Kampe et al., 2003), and in a communication task in which participants followed spoken instructions, the mere presence of a director resulted in increased activity in a number of brain regions, including the medial prefrontal cortex, bilateral superior temporal sulcus, middle temporal sulci and the extrastriate body area (Dumontheil, Küster, Apperly, & Blakemore, 2010).

Research on communication has not only looked at what happens in the brain of individual communicators. Recent years have seen an increased interest in brain-to-brain coupling during communication. For example, Stephens, Silbert, & Hasson (2010) scanned a speaker while she told an autobiographical story in the MRI scanner. They then scanned a group of listeners while they listened to this narrative, and calculated inter-subject correlations between the speaker's and the listeners' fMRI data. They found that neural activity in speakers and listeners' brains is spatially and temporally coupled. Later studies using similar paradigms have provided additional evidence for extensive inter-individual neural synchronization during communication (Silbert, Honey, Simony, Poeppel, & Hasson, 2014; Spiegelhalder et al., 2014).

**BOX 1.1: Mentalizing**

Mentalizing refers to people's ability to mentally put themselves in other people's shoes, i.e. tracking and understanding what other people think or feel. This process can be both explicit and implicit (Frith & Frith, 2012). The terms *mentalizing* and *theory of mind* are often used interchangeably, although theory of mind suggests a more conscious process (Frith & Frith, 2012). There is considerable debate about how and when children's theory of mind skills develop, as well as about whether this skill is uniquely human (Perner & Ruffman, 2005).

The ability to track another person's beliefs and intentions has been tested using a variety of tasks, including by asking participants to judge false belief stories, to answer questions about social animations and to view photographs of eyes (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Van Overwalle, 2009). Neuroimaging studies of mentalizing have consistently found increased activity in a series of brain regions that include the medial prefrontal cortex, temporoparietal junctions and precuneus (Amodio & Frith, 2006; Schurz et al., 2014; Van Overwalle, 2009).

**Some notes on methods**

It is challenging to study communicative language processing in the lab. The quintessential context in which people speak and listen, dialog, is interactive, free-flowing, self-generated and multimodal. These characteristics pose serious challenges for researchers studying communicative language processing. On the one hand, you aim to achieve ecological validity, making sure that you do not lose the key features and complexity inherent to conversation. On the other hand, you need to find ways to measure and quantify the aspect of communicative language processing you are interested in. This involves imposing a certain level of experimental control over the interaction, and often requires repetitions of the phenomenon of interest over the course of the experiment. In order to balance these opposing forces, researchers inevitably have to compromise (Kuhlen, Allefeld, Anders, & Haynes, 2015). A wide variety of solutions has been developed, including imposing task structure on the interaction by using referential communication games, using prerecorded naturalistic stimuli, limiting communicative means and using confederates instead of naive participants (Kuhlen et al., 2015).

Neuroimaging studies of communicative language production face additional challenges (Kuhlen et al., 2015). One of the challenges of fMRI is the low temporal resolution of the method (see Box 3), which makes it difficult to tease apart the neural response to events that

occur close in time. This constraint typically requires social interactions in the scanner to be broken down into distinguishable events, thus restricting the interactive and free-flowing nature of the interactions. Another challenge for fMRI experiments on communication is that speaking in the scanner can introduce various stimulus-correlated artifacts. Head motion in particular can cause poor realignment of the functional images. In addition, stimulus-correlated breathing can cause artifacts in the fMRI data during speaking (Farthing et al., 2007). Finally, a straightforward limitation of using fMRI for research on communication is that in-person face-to-face verbal interactions are impossible when one of the interlocutors lies in a dark and noisy MRI scanner.

In sum, studying communicative language production using fMRI requires researchers to find an optimal balance between experimental control and ecological validity, while considering the limitations of fMRI. One of the main decisions I made for the experiments in this thesis is to test pairs of naïve participants, because confederates can affect the naturalness of the interaction (see Kuhlen & Brennan, 2012 for a more detailed discussion of the use of confederates). In addition, I imposed a number of constraints on the experimental tasks I used to achieve an optimal level of experimental control, for example, by only allowing one-directional communication (chapters 2 and 3), restricting free-flowing interactions to a pre-scanning session (chapter 4), or limiting participants' communicative means (chapter 5).

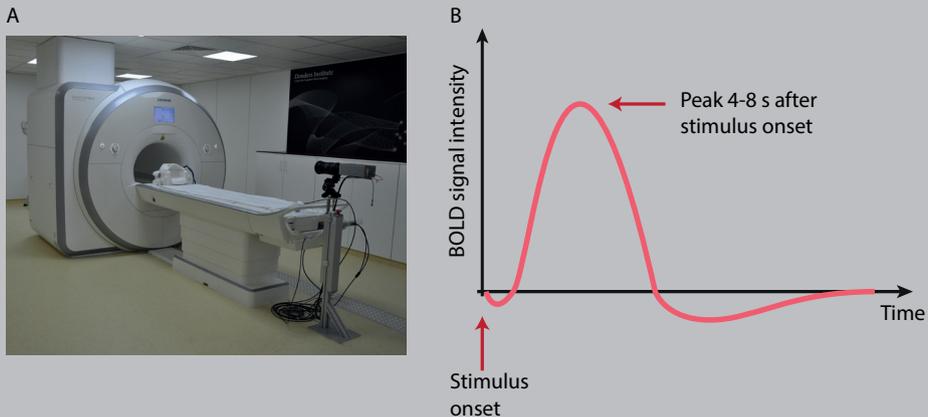
**BOX 1.2: Eye-tracking and visual world paradigms**

Moving your eyes allows you to see part of the visual world around you in fine detail. Most of the time, the part of the world you are looking at is also the part you are focusing your attention on. Eye-tracking thus provides a measure of visual attention, and has been widely used to study visual perception, reading and language processing. Cooper (1974) was the first to find a systematic relationship between eye movements and speech processing. He showed that listeners tend to spontaneously look at pictures that are semantically related to the spoken narratives they hear and that listeners' eye movements to these semantically related pictures are closely time-locked to the spoken narratives.

The time-locked relationship between speech processing and visual attention later formed the basis for visual world paradigms (Huettig, Rommers, & Meyer, 2011). In these paradigms, participants typically listen to prerecorded speech while viewing an experimental display containing a set of objects or words or a picture of a scene. The speech often contains words that are phonologically, semantically or visually related to the words or pictures in the display, allowing researchers to study how different types of linguistic information are integrated with the visual environment. In a line of similar production research, participants are asked to describe scenes or sets of objects while their eye gaze is tracked. This paradigm allows researchers to investigate how speakers generate utterance plans and coordinate planning and articulation (Huettig et al. 2011). More recently, both these production and comprehension paradigms have been adapted to study the role of common ground in interactive language processing (e.g., Hanna, Tanenhaus, & Trueswell, 2003; Keysar, Barr, Balin, & Brauner, 2000; Brown-Schmidt & Tanenhaus, 2008). In these studies, participants often see an array containing objects, some of which are occluded from their interlocutor's point of view. Participants are instructed to produce a sentence or to listen to a sentence about one of the objects in the array. On critical trials, they need to consider the distinction between common and privileged information in order to correctly interpret or unambiguously describe the target object. The high temporal resolution of eye-tracking makes it a very suitable and relatively unobtrusive method to study when common ground influences language processing.

**BOX 1.3: fMRI**

The main method used in this thesis is functional magnetic resonance imaging (fMRI), which measures the Blood Oxygenation Dependent (BOLD) signal. The BOLD signal is measured while people perform a cognitive task in the MRI scanner. When there is increased neuronal activity in a brain area in response to the task, this brain area will consume more oxygen and will receive an influx of oxygenated blood to compensate for the increased oxygen consumption. These changes in the blood oxygenation level can be measured using fMRI, as oxygen-rich blood and oxygen-poor blood have different magnetic properties. fMRI is therefore an indirect measure of brain activity: we measure hemodynamic changes in response to neuronal activity rather than the neuronal activity itself. The blood flow changes occur several seconds after the onset of neuronal activity. The BOLD signal therefore always lags behind on the actual neuronal activity, as shown in Figure 1.1B. For this reason, the temporal resolution of fMRI is in the order of seconds, making it unsuitable to answer questions about the exact timing of cognitive processes. However, the spatial resolution of the technique is very good, making fMRI an excellent and non-invasive technique to localize the brain areas that are involved in specific cognitive functions.



**Figure 1.1:** An MRI scanner (A) and the BOLD response (B). The BOLD response only peaks 4-8 seconds after stimulus onset, giving the method poor temporal resolution.

## Objectives and outline of the thesis

The overarching goal of this thesis is to investigate the cognitive and neural mechanisms underlying communicative language production. The first objective is to investigate when and how common and privileged ground information affect speech planning in a moment-to-moment fashion. The second objective is to study how the process of grounding affects the neural representations of jointly established linguistic labels. Most of the existing neuroimaging research on communication has compared communicative and non-communicative settings without considering the more fine-grained distinctions and processes identified in the psycholinguistic literature on this topic. In this thesis, I therefore also aim to link these two lines of literature by combining existing psycholinguistic paradigms with eye-tracking and fMRI.

In **chapter 2**, I investigated how and when common ground affects speech planning by tracking speakers' eye movements while they played a referential communication game. In **chapter 3**, I used a similar paradigm in an fMRI experiment in which I manipulated whether speakers had to distinguish between common and privileged ground in order to communicate efficiently with an addressee, in addition to comparing language production in a communicative and a non-communicative context. **Chapter 4** shifts the focus to the role of communication and collaboration in memory encoding and retrieval. In this experiment, participants learned labels in collaborative and individual contexts and later retrieved these labels in the MRI scanner. In **chapter 5**, I followed up on this study by investigating how the process of coordinating meaning shapes the neural representations of jointly established linguistic labels. Finally, in **chapter 6**, I discuss the findings of this thesis in relation to the existing literature and highlight interesting avenues for future research.



An early influence of common  
ground during speech planning



## **Abstract**

In order to communicate successfully, speakers have to take into account which information they share with their addressee, i.e. common ground. In the current experiment we investigated how and when common ground affects speech planning by tracking speakers' eye movements while they played a referential communication game. We found evidence that common ground exerts an early, but incomplete effect on speech planning. In addition, we did not find longer planning times when speakers had to take common ground into account, suggesting that taking common ground into account is not necessarily an effortful process. Common ground information thus appears to act as a partial constraint on language production that is integrated flexibly and efficiently in the speech planning process.

Based on:

Vanlangendonck, F., Willems, R. M., Menenti, L., & Hagoort, P. (2016). An early influence of common ground during speech planning. *Language, Cognition and Neuroscience*, 31(6), 741-750.

## Introduction

A key question in language production is when and how speakers take into account which knowledge is shared between speaker and addressee (*common ground*) and which information is only available to the speaker (*privileged ground*) (Clark & Marshall, 1978). According to Grice's Maxim of Quantity, speakers should make their contributions as informative as is required for the current purpose of an interaction (Grice, 1975). Information that is not shared between interlocutors and that is uninformative should therefore not influence the language production process. However, speakers do not always successfully ignore privileged information. Imagine a situation in which a speaker sees two bottles of different sizes, the smallest of which is occluded from their addressee's point of view. If the speaker asks the addressee to hand him the mutually visible bottle, he can either describe it using a size adjective (*the large bottle*) or a bare noun (*the bottle*). According to Grice's Maxim of Quantity, the speaker should not use a size-contrasting adjective, because the smaller bottle is occluded from the listener's point of view. However, speakers in this situation regularly use a size adjective (Horton & Keysar, 1996; Wardlow Lane & Ferreira, 2008; Yoon, Koh & Brown-Schmidt, 2012). In this case, an overspecification does not prevent the listener from identifying the intended referent. However, speakers also overspecify referring expressions when an overspecified expression can cause a misunderstanding (Wardlow Lane & Ferreira, 2008), and when they are instructed to conceal privileged information (Wardlow Lane, Groisman & Ferreira, 2006).

The question why speakers fail to completely ignore privileged information is still under debate, and two main theories have emerged to explain these results. According to Constraint-Based Processing models, speakers keep their addressee in mind from the earliest stages of utterance planning in a probabilistic, constraint-based way, resulting in early effects of common ground (Brennan & Hanna, 2009; Hanna, Tanenhaus & Trueswell, 2003; Horton & Gerrig, 2002; Tanenhaus & Trueswell, 1995). In this view, privileged and common ground information act as partial constraints for language processing. When privileged information is salient, as in the example with the additional bottle above, speakers may fail to ignore it. In contrast, the Monitoring and Adjustment theory proposes that keeping track of the distinction between shared and privileged information during utterance planning is resource-intensive and often not necessary (Horton & Keysar, 1996). According to this view, speakers initially design their utterances from their own egocentric perspective, and common ground only comes into play at a later stage when speakers monitor their utterance (Keysar, Barr & Horton, 1998). This model thus predicts that speakers initially do not distinguish between privileged and common ground information during utterance planning. Monitoring and adjusting the initial utterance plan is considered effortful and requires additional planning time (Horton & Keysar, 1996). In this view, speakers fail to ignore privileged information when their initial egocentric plan is not corrected.

In the current experiment, we investigated when and how shared and privileged visual information affect the production of referring expressions. Both the Constraint-Based Processing models and the Monitoring and Adjustment theory predict that privileged information *can* affect the production of referring expressions, but they make different predictions about *when* speakers take into account the distinction between shared and privileged ground. Constraint-Based Processing Models predict early effects of common ground, although privileged information may not always be successfully ignored. The Monitoring and Adjustment theory predicts effects of common ground only after an initial egocentric planning stage. Previous studies have mostly focused on speakers' adjective use. In the current experiment, we also measured eye-tracking data and planning durations in order to gain more insight into the timing of the underlying processes. We used the high temporal resolution of eye-tracking to test whether speakers distinguish between shared and privileged information during the early stages of planning a referring expression. In addition, we investigated whether ignoring privileged information during language production is resource-intensive as predicted by the Monitoring and Adjustment model by measuring speech planning durations. If speakers monitor and adapt their initial egocentric utterance plans when they design referring expressions that take into account their addressee's perspective, this should result in longer planning durations. Finally, we were interested in whether speakers avoid egocentric descriptions when those threaten communicative success. We therefore introduced two types of conditions. In the advisable audience design condition, a failure to ignore privileged information was overinformative, but the intended referent was clear. In the obligatory audience design condition, failing to ignore privileged information made it impossible for the listener to infer the intended referent. This condition, used here for the first time, thus provides a stronger test for speakers' sensitivity to communicative success than the advisable audience design condition.

## Method

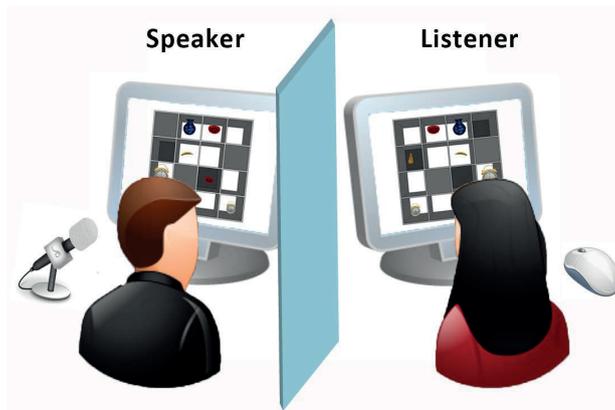
### Participants

22 pairs of native Dutch speakers played a computerised version of a referential communication game (Keysar, Barr, Balin & Brauner, 2000; Yoon, Koh & Brown-Schmidt, 2012). Participants did not know each other before the start of the experiment. Data from one pair were excluded from the analyses due to experimenter error, and data from another pair were excluded because the speaker's responses could not be coded. The remaining participants consisted of 6 men and 34 women (mean age: 21.5 years, range 18-29). Half of the eye-tracking data from one pair were lost due to equipment malfunction. The remaining eye-tracking data from this pair were included in the final dataset.

## Procedure

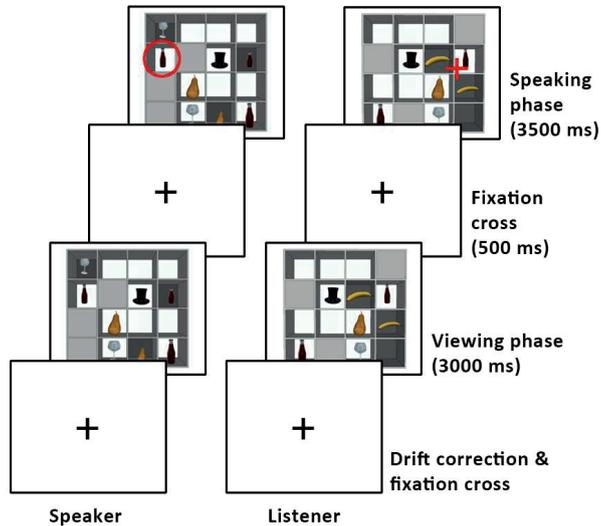
A coin toss was used to randomly assign participants to the roles of speaker and listener for the duration of the experiment. The speaker and listener were seated at separate monitors, separated by a screen to prevent them from seeing each other's monitor. Throughout the experiment, participants saw opposite sides of a 4 x 4 array containing objects of different sizes (Figure 2.1). Each array contained 3 closed slots on each player's side, allowing us to manipulate which objects were in common ground (i.e. visible to both participants).

Participants completed 288 trials in total. On each trial, the speaker described a specific object in the array in a way that would allow the listener to select the correct object from the array. During the first phase of the trial (3000 ms), the speaker and listener each saw their side of the array. Then the speaker was cued by means of a red circle around one of the objects, and described this object for the addressee (Figure 2.2). On the basis of the speaker's description, the listener selected an object by means of a mouse click. The cued object was always mutually visible. Speakers were instructed not to use descriptions referring to the position of the object in the array, such as 'upper left corner' or 'rightmost'. Participants successfully refrained from using such spatial descriptions (see results section). In addition, they were instructed not to ask questions or give each other feedback during the experiment. Speakers were not given any on-screen or verbal feedback about the listener's performance during the experiment. Participants were seated next to each other in the lab, so speakers could occasionally hear the listeners' mouse clicks.



**Figure 2.1:** Overview of the set-up. Speakers and listeners viewed opposite sides of 4 x 4 arrays containing objects. Some objects were only visible to one of the participants, allowing us to manipulate which objects were in common ground. Speakers' responses and eye movements were recorded.

Before starting the experiment, participants practised the task together using a real array and real objects. They jointly placed the objects in the array, viewed each other's perspective, and performed the task. Then participants practised the task on the computer; during part of these test trials they were allowed to see both monitors and they could give feedback to each other.



**Figure 2.2:** Trial sequence from the speaker's and the listener's point of view. In the first phase of each trial, speakers and listeners each viewed their side of the array. In the second phase of the trial, the speaker described the object that was indicated with a red circle, and the listener selected an object using the cursor.

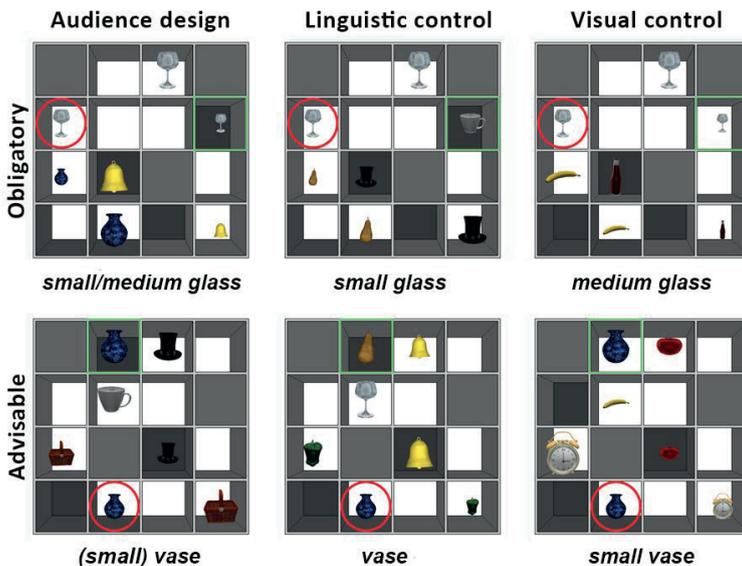
Speakers' descriptions were recorded and their eye movements were tracked using a head-mounted EyeLink II eye-tracker (SR Research) with a 250 Hz sampling frequency. We tracked both pupil and corneal reflection whenever possible. The distance between the speaker and the monitor was approximately 57 cm. We calibrated the eye-tracker at the start of the experiment, after a break halfway through the experiment and whenever deemed necessary. Drift correction was performed before each trial.

## Materials

We manipulated the number, size and visibility of the relevant objects to create 6 conditions (Figure 2.3). In the audience design conditions (left column Figure 2.3), speakers saw an extra competitor object that the listener could not see. In the *obligatory* audience design condition, speakers saw 3 relevant identical objects of different sizes: one target object, one occluded competitor object and one mutually visible object. The target object was always the medium-sized object of the 3 objects. If speakers described this object from their own perspective, they would call it the *medium* object. In this case, their addressee had a 50 per

cent chance of selecting the wrong object. On the other hand, if speakers considered the perspective of their addressee, they would ignore the occluded object and call the medium-sized object *small* or *large*. In the *advisable* audience design condition, speakers saw 2 relevant identical objects of different sizes: one target object and one occluded competitor object. Given that their addressee could see the target object but not the competitor object, speakers did not have to use a contrasting size adjective. However, unlike in the obligatory audience design condition, listeners were still able to select the correct object if the speaker did not consider the perspective difference.

We created two types of control conditions in which speakers and listeners saw the same number of relevant objects. In the linguistic control conditions (middle column Figure 2.3), the occluded object was replaced by another, unrelated object. As a result, speakers saw one relevant object fewer in these conditions than in the audience design conditions. We called these *linguistic* control conditions, because speakers were expected to produce the same description on these trials as on successful trials in the audience design conditions. In the visual control conditions (right column Figure 2.3), the object that was occluded in the audience design conditions was visible to both participants.



**Figure 2.3:** Overview of the 6 conditions from the speaker's point of view, and the expected speaker responses. In the audience design conditions, speakers can either describe the target objects (circled in red) from their own perspective ("medium glass", "small vase"), or take into account their addressee's perspective ("small glass", "vase"). There is no relevant perspective difference in the linguistic and visual control conditions. The green squares were added to the figure for clarification, and were not visible to the participants during the experiment. They indicate the objects that differ between conditions.

As a result, speakers and listeners could both see all relevant objects. We called these *visual control conditions*, because speakers see the same number of relevant objects as in the audience design conditions. Neither of these control conditions required the speaker to take into account the perspective difference with their addressee in order to communicate successfully.

We created 12 different empty virtual arrays. The arrays were filled with 6 to 8 objects chosen from a total of 22 objects. Objects were selected from the Object Databank (courtesy of the Michael Tarr lab, Brown University, Providence, RI). Each object could appear in 4 different sizes to make sure that participants could not rely on absolute size. Depending on the condition, speakers saw 1, 2 or 3 relevant objects of the same type but of different sizes. The remaining objects were fillers that also appeared in sets of 1, 2 or 3 objects of the same type to make sure participants could not predict which objects would be relevant. We made sure that the speaker and the listener always saw the same total number of objects in a trial by adding additional filler objects to the occluded slots if needed. We have created a visual overview depicting the trial construction, which can be found in the supplementary materials. For each speaker, we created 288 trials by constructing 48 triplets of obligatory trials and 48 triplets of advisable trials. Each triplet consisted of one audience design trial, one linguistic control trial and one visual control trial. Figure 2.3 shows examples of one obligatory triplet (top row) and one advisable triplet (bottom row). The target and competitor objects per triplet were identical in type, size and location, with the exception of the occluded object in the linguistic control condition. The number, proportion and location of filler objects were kept constant within each triplet, but we varied the type and the size of the filler objects in order to avoid memory effects. As can be seen in Figure 2.3, the audience design and the linguistic control trials in a triplet were created using the same array. The array used in the visual control condition differed from this array in the location of one closed slot. Each of the twelve virtual arrays was used to create four obligatory and four advisable triplets of trials. The only thing these triplets had in common was the array that was used as a starting point. Target object-location pairings were only repeated within a triplet, i.e. three times throughout the experiment.

During the experiment, trials were presented in blocks of six trials that were created using the same array. Each block contained no more than one trial from a triplet. The order of the trials within each block was randomized, and we randomized the blocks of trials so that neighboring blocks did not use the same array. We created unique stimulus lists for each participant, detailing which object should appear where in each array. Based on these lists, the stimulus pictures, as presented to the participants, were assembled online in Presentation software.

## Data analysis

We coded the sound files offline for adjective use. For the obligatory trials, we coded the use of *small/large* and *medium* adjectives (*klein/groot* and *middelgroot* in Dutch). For the advisable trials, we coded the use of bare nouns and *small/large* responses. In addition, we coded errors, false starts, repairs, speech unrelated to the task, and responses exceeding the response interval of 3500 ms, and removed trials that contained any of these from further analysis (10.45% of trials). Listener performance was calculated by determining whether listeners clicked in the right slot within the 3500ms-response interval. Finally, we computed planning durations (from cue onset until speech onset) and speaking durations (from speech onset until speech offset) from the sound files using a custom-made Matlab script. The script filtered and smoothed the waveform, determined an initial speech onset and offset by looking for sustained periods of silence and speech in the spectrogram and then calculated a more precise speech onset and offset by using increasingly smaller windows and lower thresholds. Informal inspection of the speech waveforms revealed that these calculations were accurate (example waveforms are included in the supplementary materials).

The statistical analyses were performed in R version 3.0.3 (R Core Team, 2014). We used the `lmer` function from the `lme4` package version\_0.999999-4 (Bates, Mächler, Bolker & Walker, 2015) to fit the linear mixed models and we built up the models from a simple to a more complex model. The linear mixed models (Baayen, Davidson & Bates, 2008) included a random intercept for subjects, a by-subjects random slope for condition, a random intercept for target object indicating which of the 22 possible objects was used as target object, and condition and number of objects as fixed effects. Models were estimated using maximum likelihood (ML) estimation. The factor *condition* had three levels (audience design, linguistic control and visual control) and the factor *number of objects* had three levels (6, 7 or 8 objects). We used the same model in every analysis, except that the reference distribution was changed depending on the dependent measure: Poisson for count data (number of fixations per trial), binomial for binary data (adjective use) and Gaussian for continuous data (planning durations, speaking durations, and fixation durations). Obligatory and advisable data were analysed separately. The obligatory conditions contained more relevant objects than the advisable conditions, so we could not compare them directly.

The main effect of condition was tested by comparing a complete model containing all random and fixed effects to a model in which the fixed effect of condition was removed (the by-subjects random slope for condition remained in the model). For each dependent measure, two planned contrasts were performed in order to compare the audience design condition with each of the control conditions. We used the `glht` function from the `multcomp` package version 1.3-2 (Hothorn, Bretz & Westfall, 2008) to run these planned comparisons. More information about the data analysis can be found in the supplementary materials.

## Results

### Adjective use

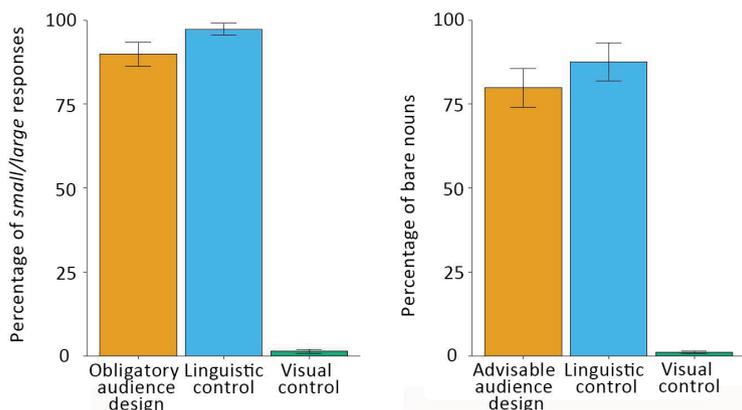
We computed the percentage of each type of response for the different conditions to find out how often speakers took their addressee's perspective into account when producing a referring expression. For the obligatory trials, we treated the presence of the adjectives *small* or *large* as opposed to *medium* as a binary dependent variable; for the advisable conditions, we treated the presence of a bare noun as opposed to a size adjective (*small* or *large*) in the speakers' descriptions as a binary dependent variable.

Speakers adapted their language use to their addressee's perspective in the majority of audience design trials (Figure 2.4), yet they failed to ignore privileged information on all trials. In the obligatory audience design condition, speakers mainly produced *small/large* responses (89.88%), although they also used *medium* responses (10.12%). *Medium* responses are accurate descriptions of the target object from the speakers' perspective, but do not allow the listener to uniquely identify the intended referent. As expected, speakers mostly produced utterances that contained a *small/large* size adjective (97.32%) in the linguistic control condition and mainly *medium* responses (98.59%) in the visual control condition. We found a significant main effect of condition on the use of *small/large* size adjectives ( $\chi^2(2) = 41.71, p < 0.001$ ). Both the difference in adjective use between the obligatory audience design condition and the linguistic control condition ( $b = 3.75, SE = 0.51, p < 0.001$ ) and the difference in adjective use between the obligatory audience design condition and the visual control condition were significant ( $b = -8.77, SE = 0.72, p < 0.001$ ).

In the advisable audience design condition, speakers generally took their addressee's perspective into account. They mostly produced bare nouns (79.84%), although they also produced *small/large* responses (20.16%). In the linguistic control condition, speakers mainly produced bare nouns (87.62%), and in the visual control condition they predominantly used *small/large* responses (98.90%). We found a significant main effect of condition on the use of bare nouns ( $\chi^2(2) = 39.88, p < 0.001$ ). Both the difference between the advisable audience design condition and the linguistic control condition ( $b = 1.64, SE = 0.28, p < 0.001$ ) and the difference between the advisable audience design condition and the visual control condition were significant ( $b = -6.60, SE = 0.60, p < 0.001$ ).

Overall listener accuracy was high: listeners selected the intended object on 93.60% of the obligatory trials and 97.70% on the advisable trials. As a manipulation check we looked specifically at listeners' performance on audience design trials on which speakers described the target object without taking their addressee's perspective into account (i.e., saying *medium* in the obligatory audience design trials and using a size adjective in the advisable

audience design trials). This was a small subset of trials, unequally divided over participant pairs (obligatory data: 2.98% of trials, advisable data: 6.14% of trials). On obligatory audience design trials, listeners only selected the correct object in 36% of these ambiguous trials; on advisable audience design trials, listeners were successful in 98.15% of trials. This confirms that our advisable/obligatory manipulation worked as expected: not taking the addressee's perspective into account forced the listener to guess the referent in the obligatory audience design trials, but not in the advisable audience design trials.



**Figure 2.4:** The percentages of speakers' small/large and bare noun responses for the obligatory (left) and advisable data (right). Bars indicate standard error of the mean. Speakers were expected to produce such responses in the audience design conditions if they took their addressee's perspective into account.

## Planning and speaking durations

We computed speakers' planning durations (from cue onset to speech onset) based on the sound recordings to test whether ignoring privileged information during utterance planning is resource-intensive. We found no significant main effect of condition in both the obligatory data ( $\chi^2(2) = 1.82, p = 0.40$ ) and the advisable data ( $\chi^2(2) = 0.84, p = 0.66$ ). Figures of the mean planning durations per condition are included in the supplementary materials.

We also calculated speaking durations (from speech onset to speech offset). If ignoring privileged information is a resource-intensive process, this might also lead to longer speaking durations. We found a significant main effect of condition on speaking durations, both for the obligatory ( $\chi^2(2) = 26.23, p < 0.001$ ) and the advisable data ( $\chi^2(2) = 45.65, p < 0.001$ ). No significant difference in speaking time was found between the obligatory audience design condition and the linguistic control condition ( $b = -17.76, SE = 8.39, p = 0.067$ ). Speaking durations were shorter in the obligatory audience design condition than in the visual control condition ( $b = 86.54, SE = 12.38, p < 0.001$ ), most likely because the

Dutch word *middelgroot* has more syllables than the other size adjectives (*klein* and *groot*). We found significant differences in speaking duration between the advisable audience design and the linguistic control condition ( $b = -26.61$ ,  $SE = 8.96$ ,  $p < 0.01$ ) and between the advisable audience design condition and the visual control condition ( $b = 224.43$ ,  $SE = 17.52$ ,  $p < 0.001$ ). However, when we removed the advisable audience design trials in which the speaker did not take the addressee's perspective into account (i.e. the speaker produced a size adjective), the difference between the advisable audience design condition and the linguistic control condition disappeared ( $b = 15.95$ ,  $SE = 14.34$ ,  $p = 0.46$ ). Figures of the mean speaking durations per condition are included in the supplementary materials.

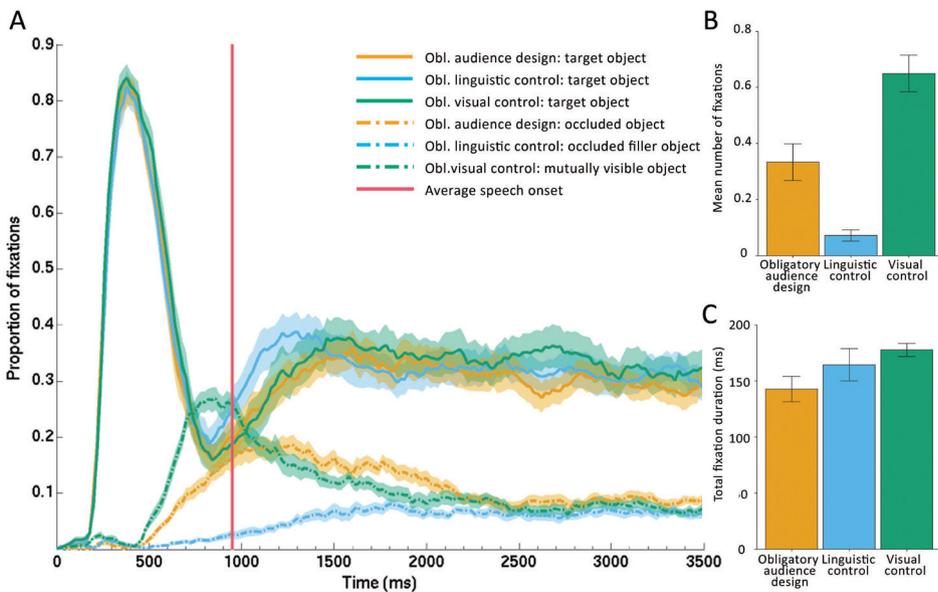
### Number of fixations per trial

In order to measure when privileged information affects the production of referring expressions, we computed the mean number of fixations on the occluded competitor object in the audience design conditions and on the objects in the same locations in the associated control conditions (the objects in the green squares in Figure 2.3). We focused our analysis specifically on the planning duration window, i.e. in each trial, the time between the moment the array with the cued object was presented and the time the speaker started talking. For both the obligatory and the advisable data, we found that speakers fixated an occluded competitor object (audience design conditions) more than an unrelated control object (linguistic control), but less than a mutually visible competitor object (visual control). Figures 2.5 and 2.6 show how the proportion of fixations on the target and occluded object changes over time and give an overview of the number of fixations per trial. Figures 2.5A and 2.6A clearly show that speakers initially fixate the target object they have to describe, and then consider the other objects in the array.

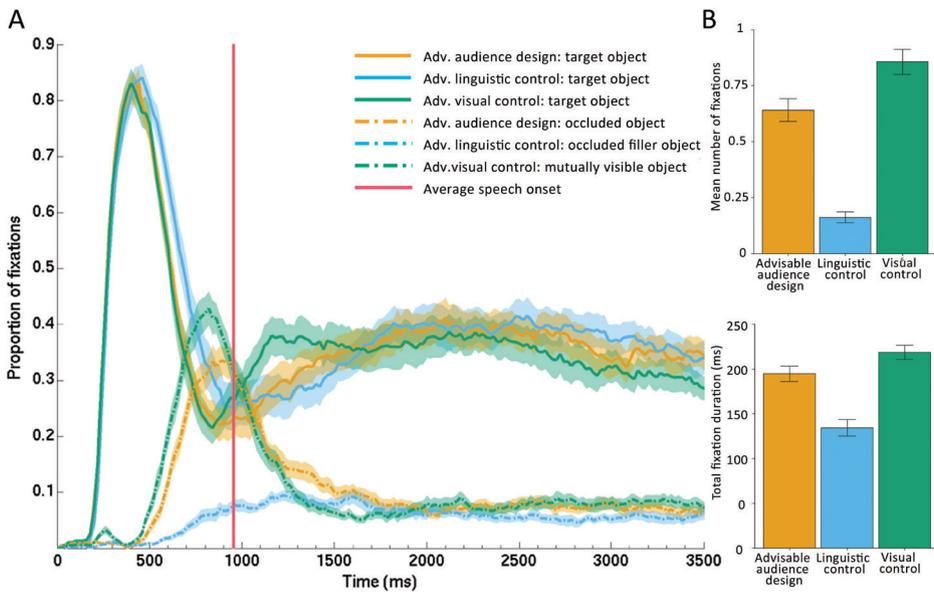
We found a significant main effect of condition on the number of fixations per trial in the obligatory data, ( $\chi^2(2) = 52.28$ ,  $p < 0.001$ ). We found significantly more fixations per trial on the occluded competitor object in the obligatory audience design condition compared to the occluded control object in the linguistic control condition ( $b = -1.57$ ,  $SE = 0.15$ ,  $p < 0.001$ ). We found significantly less fixations per trial on the occluded competitor object in the obligatory audience design condition compared to the mutually visible object in the visual control condition ( $b = 0.89$ ,  $SE = 0.12$ ,  $p < 0.001$ ). To rule out that these differences were driven by the small percentage of obligatory control trials on which speakers responded with a *medium* response, we ran the same analysis including only trials in which speakers' responses took into account the listener's perspective (i.e., a *medium* response in the visual control condition, a *small/large* response in the audience design and linguistic control conditions). The differences remained significant in this analysis. Across the obligatory conditions, the mean number of fixations per trial is relatively low, even when the competitor

object was relevant (visual control condition). Some participants reported that they paid special attention to objects that appeared in triplets during the first half of the trials, which may have allowed them to plan their utterance after the cue without additional fixations on the other objects.

We observed a similar pattern in the advisable data. We found a significant main effect of condition on the number of fixations per trial, ( $\chi^2(2) = 41.99, p < 0.001$ ). We found a significant difference in the mean number of fixations per trial between the advisable audience design condition and the linguistic control condition ( $b = -1.44, SE = 0.11, p < 0.001$ ) and between the advisable audience design condition and the visual control condition ( $b = 0.31, SE = 0.063, p < 0.001$ ). This pattern remained even when we only included the trials in which speakers' responses took into account the listener's perspective (i.e., an adjective in the visual control condition, a bare noun in the advisable audience design and linguistic control conditions). These results show that speakers already distinguish between common and privileged ground while planning their utterance, although they do not completely ignore privileged information.



**Figure 2.5:** Eye-tracking results for the obligatory data. (A) Proportions of speakers' fixations to the target and occluded objects from the moment speakers were cued until the end of the trial. Solid lines indicate fixations to the target object; dashed lines represent fixations to the occluded competitor object, the occluded competitor object or the mutually visible competitor object. Bounded lines indicate the standard errors. (B) Mean number of fixations (C) and total fixation time during the planning duration window on the occluded competitor object in the obligatory audience design condition, on the occluded filler object in the linguistic control condition and on the mutually visible competitor object in the visual control condition.



**Figure 2.6:** Eye-tracking results for the advisable data. (A) Proportions of speakers' fixations to the target and occluded objects from the moment speakers were cued until the end of the trial. Solid lines indicate fixations to the target object; dashed lines represent fixations to the occluded competitor object, the occluded filler object or the mutually visible competitor object. Bounded lines indicate the standard errors. (B) Mean number of fixations (C) and total fixation time during the planning duration window on the occluded competitor object in the advisable audience design condition, on the occluded filler object in the linguistic control condition and on the mutually visible competitor object in the visual control condition.

### Total fixation duration per trial

We computed the total duration of fixations on the occluded competitor object in the audience design conditions and on the objects in the same locations in the associated control conditions. We again limited the analysis to fixations during the planning duration window. Fixations that exceeded the planning duration window were truncated and we treated trials without fixations on the occluded object as missing data. Speakers fixated the unrelated occluded object in the linguistic control condition in only 6.49% of obligatory linguistic control condition trials, and on 16.08% of advisable linguistic control trials. Given that there were so few of these trials, we did not contrast the audience design conditions and the linguistic control conditions for this dependent variable. An overview of the average total fixation durations can be found in Figures 2.5C and 2.6C.

Speakers looked longer at mutually visible objects (visual control conditions) than at objects that were occluded from the listener's point of view (audience design conditions). We found significant main effects of condition in both the obligatory ( $\chi^2(2) = 14.15, p < 0.001$ ) and the advisable data ( $\chi^2(2) = 27.12, p < 0.001$ ). We found a significant difference in total

fixation duration between the obligatory audience design condition and the visual control condition ( $b = 31.82$ ,  $SE = 9.41$ ,  $p < 0.01$ ). Similarly, speakers spent less time looking at the occluded competitor object in the advisable audience design condition than at the mutually visible object in the visual control condition ( $b = 22.53$ ,  $SE = 7.16$ ,  $p < 0.01$ ). These results again support the finding that speakers already treat common and privileged information differently while planning their utterance.

## Discussion

In this experiment, we investigated the effect of privileged information on the production of referring expressions. Speakers mainly produced referring expressions that took into account their addressee's visual perspective. However, the availability of privileged information also led to the production of utterances that did not take this perspective difference into account. These findings are in line with previous studies using similar paradigms (Yoon, Koh & Brown-Schmidt, 2012; Wardlow Lane, Groisman & Ferreira, 2006). Interestingly, speakers even failed to completely ignore privileged information when it harmed communication (i.e., the obligatory audience design condition). In a previous study, Wardlow Lane and Ferreira (2008) also found that speakers were not able to fully ignore privileged information even when it threatened communicative success. However, in their study target objects were in privileged ground in certain filler trials, meaning that speakers could not systematically ignore privileged ground information. In the current experiment, speakers could fully ignore occluded objects throughout the experiment, yet even threats to communicative success did not prevent occasional interference of privileged information.

In addition to studying the form of speakers' referring expressions, we collected planning durations and eye-tracking data to address when and how privileged information affects the production of referring expressions. Speakers fixated occluded competitor objects less than relevant competitor objects both in terms of the number of fixations and the total fixation duration. This suggests that speakers did not initially treat privileged information in an egocentric manner as predicted by the Monitoring and Adjustment hypothesis, but that they take into account which objects are visible to their communication partner. More specifically, speakers used the information that was available to them in the first half of each trial to distinguish between common and privileged ground when they were planning their referring expressions. Despite this early distinction, speakers did not fully ignore privileged information when planning their referring expression. We found more fixations on the occluded target objects than on unrelated occluded objects. This shows that speakers' general success at producing utterances that took into account their addressee's perspective was not the result of fully ignoring all occluded objects. Fixations to the occluded competitor object might result from speakers' occasional failure to ignore their own perspective.

Alternatively, speakers may deliberately fixate the occluded object to check whether it is in common ground, for example because they failed to look at the occluded object during the viewing phase or as part of monitoring during speech production. Based on our data, we cannot exclude either option.

To test the prediction of the Monitoring and Adjustment theory that audience design should lead to additional planning time (Horton & Keysar, 1996), we computed and compared planning durations. However, unlike Horton & Keysar (1996), we did not find any significant differences in planning duration between the audience design conditions and the control conditions. One important difference between these experiments is that in our experiment participants had the opportunity to view the array and objects before they had to describe the target object. We included the viewing phase, because in most real-life communicative settings, speakers also have the opportunity to process their environment before they start talking. The viewing phase may have facilitated the speech planning process by allowing speakers to compute which objects were in common ground before they started planning their utterances. In addition, we repeated the same 22 images scaled to different sizes during the experiment, which may have allowed speakers to learn to identify absolute object size and may have thus facilitated target object naming over the course of the experiment. However, we consider this possibility unlikely given that the objects had different sizes to begin with and could appear on screen in four possible sizes, making it hard to determine absolute object size.

Taken together, our results show that common ground does not function as a complete constraint on the production of referring expressions, but does exert an early effect during utterance planning. Speakers tried to use the information that was available to them during the first half of each trial to restrict common ground when they were planning their utterance, as reflected in the lower number of fixations on the occluded object during planning in the second half of the trials. When speakers fixated the occluded object during planning, this often did not prevent them from tailoring their referring expression for their addressee. These results support Constraint-Based Processing models. In this view, common and privileged ground act as probabilistic constraints to guide language processing in combination with other constraints such as context (Brown-Schmidt & Hanna, 2011). Given the lack of differences in planning durations and speakers' relative success at tailoring their referring expressions for their addressee, weighing the available shared and privileged information appears to be a relatively efficient process. The finding that common ground functions merely as a partial constraint is perhaps best understandable when you consider that in many communicative situations the goal of the interaction is exactly to provide privileged information (e.g. giving directions or responding to questions). Maybe speakers' attention to privileged objects should therefore be thought of as a useful feature

of communicative language processing rather than a failure to ignore irrelevant information. In addition, referring expressions that contain uninformative privileged information often do not prevent the addressee from identifying the intended referent, although they may temporarily confuse the addressee (Engelhardt, Bailey, & Ferreira, 2006).

Our findings of early and partial effects of common ground on language *production* are similar to previous findings in language *comprehension* research showing early effects of common ground in reference resolution (Hanna, Tanenhaus & Trueswell, 2003; Heller, Grodner & Tanenhaus, 2008; Nadig & Sedivy, 2002). This suggests that common ground may affect language production and language comprehension in similar ways.

## Conclusion

In line with previous studies using similar paradigms (Yoon, Koh & Brown-Schmidt, 2012; Wardlow Lane, Groisman & Ferreira, 2006), we found that speakers cannot completely ignore privileged information during language production and that the availability of privileged information can lead to the production of utterances that do not take the addressee's visual perspective into account. However, we found no evidence that adapting to your addressee's visual perspective requires additional planning time, suggesting that audience design is not necessarily effortful. Combined, these results suggest that the available privileged and shared information can efficiently constrain language production. These findings support Constraint-Based Processing models that predict that common and privileged information is incorporated into language processing in a probabilistic fashion. Our results show that common ground does not exert an all-or-nothing influence on language production. Instead, common ground influences the production process as a partial constraint already during utterance planning.

## Supplementary materials

### Trial construction

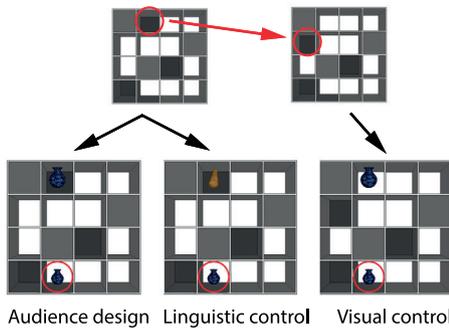
We created 12 different empty virtual arrays.



Each array differed from two other arrays in the location of one closed slot.

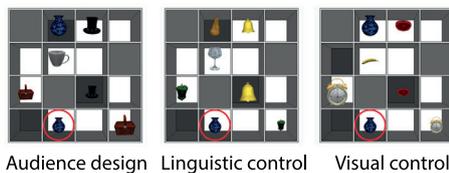


Each array was used as a starting point to create 4 obligatory and 4 advisable triplets of trials. Each triplet consisted of one audience design trial, one linguistic control trial and one visual control trial. The triplets were created using the following steps:



1 **Select an array** to be used for audience design trial and linguistic control trial. Select one of the arrays that differ in the location of one closed slot to create the visual control trial.

2 Select the type, size and location of **target and competitor objects**. Target-location pairings were repeated within but not across triplets.



3 Add **filler objects**. The number, proportion and location of filler objects were kept constant in a triplet, but the type and size of filler objects varied.

$$12 \text{ arrays} \times (4 \text{ obligatory} + 4 \text{ advisable triplets per array}) = 96 \text{ triplets} \\ = 288 \text{ trials}$$

**Supplementary Figure 2.1:** Visual overview of the trial construction.

## Additional information about the linear mixed models

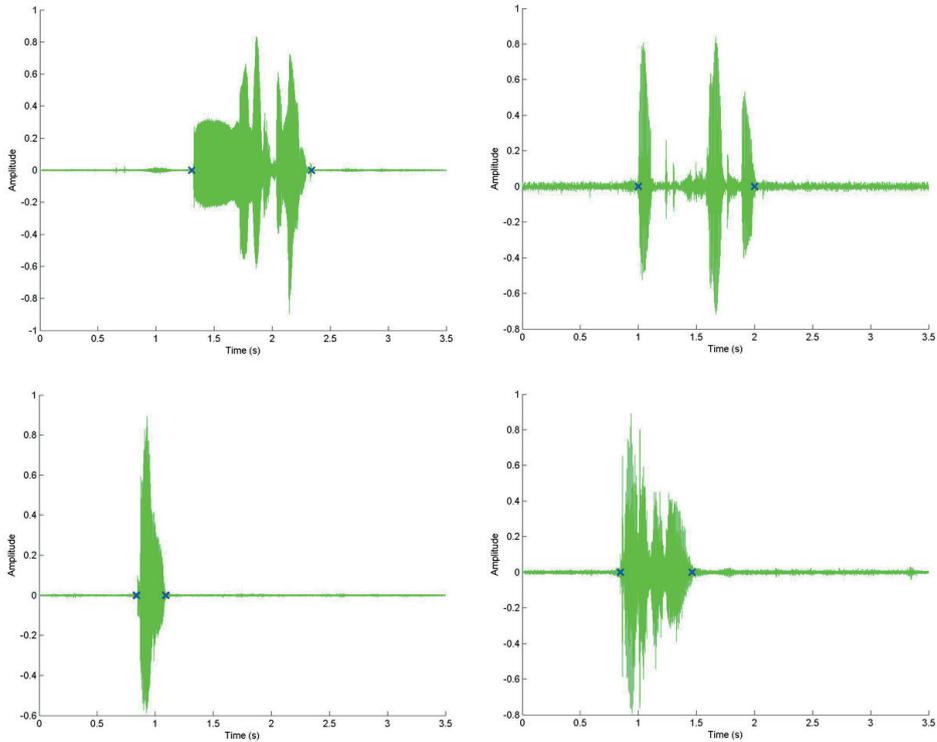
We used linear mixed models with a random intercept for subjects, a by-subjects random slope for condition, a random intercept for target object indicating which of the 22 possible objects was used as target object, and condition and number of objects as fixed effects. The factor *condition* had three levels (audience design, linguistic control, visual control) and the factor *number of objects* had three levels (6, 7 or 8 objects). We changed the coding of the predictors from the default R treatment coding to sum coding. We created separate models for the obligatory and the advisable data. Our final models looked like this:

```
example_model = lmer(EXAMPLE_DV ~ Trial_type + Number_of_objects + (1+Trial_type|Subject_nr) + (1|Object_type), DATA, REML=FALSE, family="FAMILY_TYPE")
```

- EXAMPLE\_DV = planning durations, speaking durations, adjective use, number of fixations or total fixation duration
- Trial\_type = condition with three levels (e.g., advisable audience design, advisable linguistic control, advisable visual control)
- Number\_of\_objects = total number of objects visible to the speaker per trial (6, 7 or 8)
- Subject\_nr = subject
- Object\_type = target object, which of the 22 different objects was used as target
- DATA = the dataset, either only the obligatory or only the advisable data
- FAMILY\_TYPE = Poisson for count data (number of fixations per trial), binomial for binary data (adjective use) and Gaussian for continuous data (planning durations, speaking durations, and fixation durations)

We used the `glht` function from the `multcomp` package to run the planned comparisons comparing the audience design condition to the control conditions. We changed the coding for the predictors in the following way to run these comparisons:

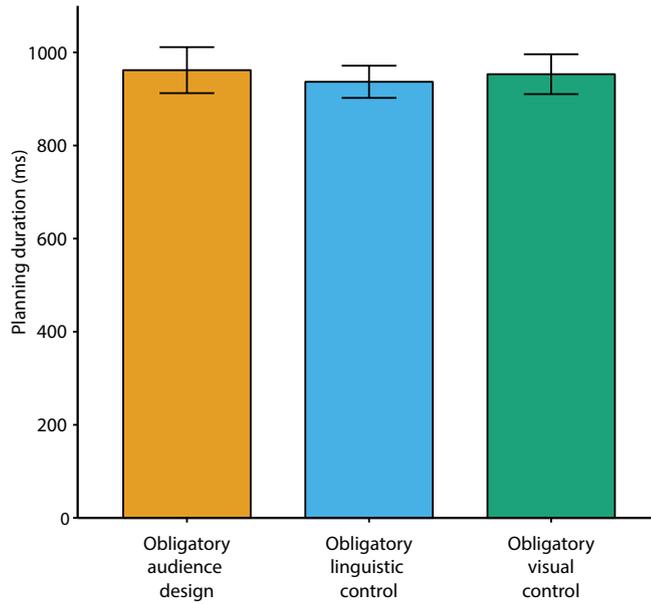
	<b>Audience design</b>	<b>Linguistic control</b>	<b>Visual control</b>
Linguistic control – audience design	-1	1	0
Visual control – audience design	-1	0	1



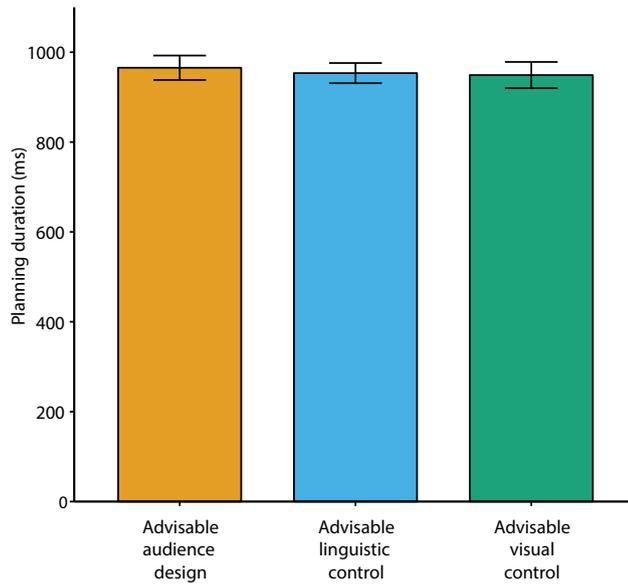
**Supplementary Figure 2.2:** Example waveforms illustrating our extraction of speech onset and offset. These example waveforms were generated by the Matlab script we used to compute the speech onset and offset of the speakers’ responses. The waveforms are plotted in green. The blue crosses indicate the speech onset and offset selected by the script. The plotted sound files were randomly selected from four different participants in the experiment. As can be seen in these examples, the script detected speech onset and offset accurately.

	Audience design	Linguistic control	Visual control
<b>Obligatory</b>	137	98	126
<b>Advisable</b>	79	29	133

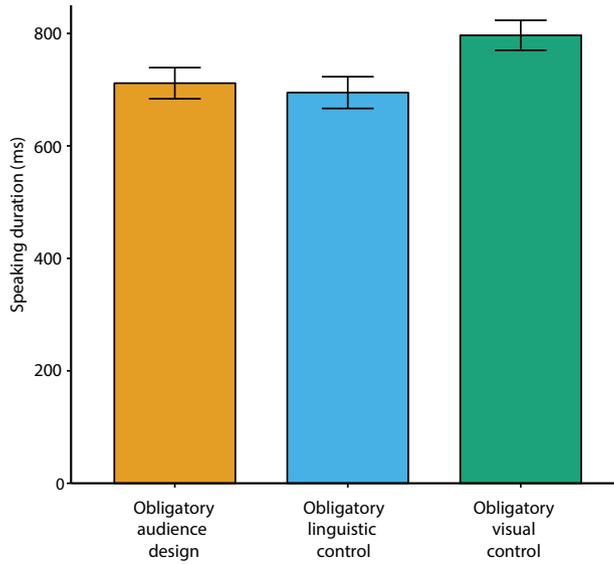
**Supplementary Table 2.1:** The number of removed trials per condition. We removed errors, false starts, repairs, speech unrelated to the task, and responses exceeding the response interval of 3500 ms. We noticed that we excluded a relatively high number trials from the advisable visual control condition (133) compared to the advisable audience design condition and the advisable linguistic control condition. Closer inspection revealed that 89 of the advisable visual control trials were removed because speakers first only mentioned the noun and then corrected themselves by adding a size adjective (e.g., “the tomato, the small tomato”).



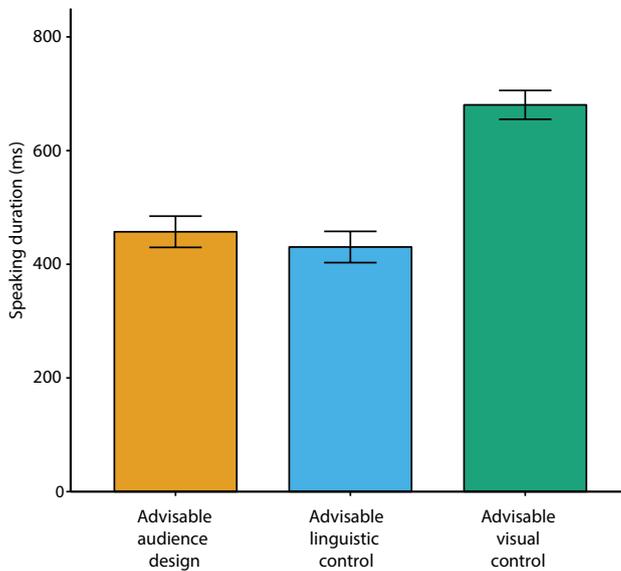
**Supplementary Figure 2.3:** Overview of mean planning durations for the obligatory conditions. Bars indicate standard error of the mean.



**Supplementary Figure 2.4:** Overview of mean planning durations for the advisable conditions. Bars indicate standard error of the mean.



**Supplementary Figure 2.5:** Overview of mean speaking durations for the obligatory conditions. Bars indicate standard error of the mean.



**Supplementary Figure 2.6:** Overview of mean speaking durations for the advisable conditions. Bars indicate standard error of the mean.





Taking common ground into account: Specifying the role of the mentalizing network in communicative language production



## **Abstract**

Several studies have shown that communicative language production as compared to non-communicative language production recruits parts of the mentalizing or theory of mind network, yet the exact role of this network in communication remains underspecified. In this study, we therefore aimed to test under what conditions the mentalizing network contributes to communicative language production. We were especially interested in distinguishing between situations in which speakers have to consider which information they do or do not share with their addressee (common vs. privileged ground information). We therefore manipulated whether speakers had to distinguish between common and privileged ground in order to communicate efficiently with the listener, in addition to comparing language production in a communicative and a non-communicative context. Participants performed a referential communicative game in the MRI-scanner as well as a similar, non-communicative task. We found that the medial prefrontal cortex, a core region of the mentalizing network, is especially sensitive to communicative contexts in which speakers have to take their addressee's needs into account in order to communicate efficiently. In addition, we found neural differences between the communicative and the non-communicative settings before speakers started to plan their utterances, suggesting that they continuously update common ground in a communicative context.

Based on:

Vanlangendonck, F., Willems, R. M., & Hagoort, P. (submitted). Taking common ground into account: Specifying the role of the mentalizing network in communicative language production.

## Introduction

Recent years have seen an increased interest in the study of the neural mechanisms supporting the social and communicative aspects of language production. A number of studies have shown that planning a communicative action as compared to a non-communicative action recruits parts of the mentalizing or theory of mind network, suggesting that people mentalize about their interlocutor's intentions and beliefs when planning a communicative action. For example, language production in a communicative task as compared to a non-communicative task was found to recruit parts of the mentalizing network: the temporoparietal junction / superior temporal sulcus (TPJ/pSTS), and the medial prefrontal cortex (mPFC; Sassa et al., 2007; Willems et al., 2010). Similarly, in a non-linguistic context, the pSTS is more activated when planning a communicative action compared to a non-communicative action (Noordzij et al., 2009) and during communicative pointing compared to non-communicative pointing (Cleret de Langavant et al., 2011).

But what sets apart communicative from non-communicative language production? In a communicative context, speakers typically have to take into account what information they share with their addressee (common ground) and what information they do not share (privileged ground). For example, when explaining the results of your latest experiment to a colleague, you should take into account how much this person knows about your research topic and whether you have talked about it before. Such adjustments require that you continuously monitor common ground to create a model of what knowledge and beliefs you have in common with your interlocutor (Clark & Krych, 2004; Clark & Marshall, 1978; Clark & Murphy, 1982). Some information for the model may be available from the start of the interaction, while other information only becomes available as the interaction unfolds (Brennan, Galati, & Kuhlen, 2010). For example, while explaining the results of your experiment to your colleague, you may discover that they know less about the topic than you anticipated, requiring you to adjust your language use accordingly. Here we extend previous studies on the neural basis of communicative language production by explicitly manipulating common and privileged ground in a language production task.

We examined the neural mechanisms that allow speakers to adapt their language use to moment-to-moment changes in common ground by manipulating which information they do or do not share with their addressee, in addition to comparing language production in a communicative and a non-communicative context. We used a well-established paradigm that allows for tight control over the linguistic utterances that are produced. Speakers in the MRI scanner either described objects to a listener outside the scanner (*communicative blocks*) or for themselves (*non-communicative blocks*), which allowed us to tap into the process of building a model of your addressee. In addition, and most importantly, within the

communicative blocks, we manipulated whether the speaker had to distinguish between common and privileged ground in order to communicate efficiently with the listener (*privileged ground vs. control conditions*). In the privileged ground condition, speakers saw additional competitor objects that were occluded from the addressee's point of view. They had to take into account that the addressee could not see these privileged-ground objects in order to communicate clearly with the addressee. In the control conditions, all relevant objects were mutually visible. In a previous eye-tracking study (Vanlangendonck, Willems, Menenti, & Hagoort, 2016), we found that speakers generally take their addressee's perspective into account in this task, although they cannot completely ignore privileged ground information.

Our hypotheses are driven by recent meta-analyses that distinguish between subfunctions in the mentalizing network (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Van Overwalle, 2009). One proposal is that the mPFC supports inferences about other people's or your own lasting psychological and social states, such as personality traits, while the TPJ is involved in inferring temporary states of other people, such as immediate goals and intentions (Van Overwalle, 2009; Van Overwalle & Baetens, 2009). In this view, building an overall model of your interlocutor during a communicative task might rely especially on the mPFC, while faster, moment-by-moment adaptations to changes in common ground could involve the TPJ/pSTS (Brennan et al., 2010). In the present experiment, we therefore hypothesized that using a mental model of your addressee might especially engage the mPFC, because the information in these models mostly relates to enduring features of the addressee. The mPFC should therefore be sensitive to the general difference between communicative and non-communicative context. On the other hand, we expected that the TPJ should be especially engaged when speakers detect a relevant perspective difference during the communicative task blocks and need to adjust their language use accordingly. That is, we expected the TPJ to be sensitive to the distinction between information that is in common ground and information that is in privileged ground.

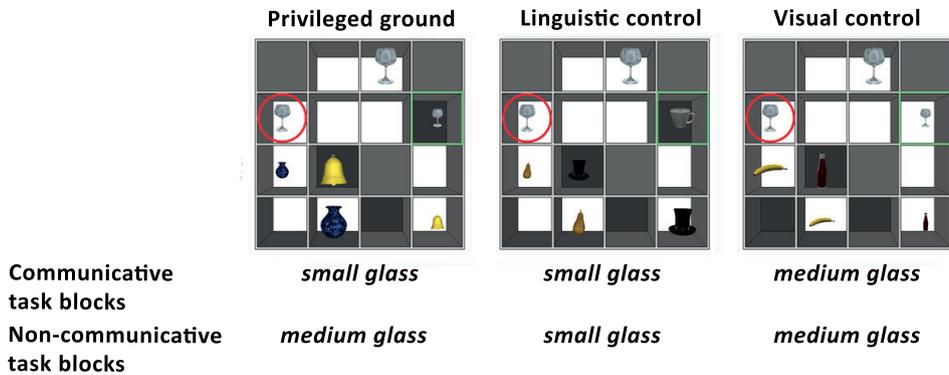
## Method

### Participants

Twenty-four pairs of right-handed native Dutch speakers participated in the study. Participants did not know each other before the start of the experiment. All participants had normal or corrected-to-normal vision and no history of neurological disease. They gave written informed consent before the start of the experiment. Data from two pairs were excluded due to technical problems and data from two additional pairs were excluded due to excessive movement by the subject in the scanner. The results of the remaining forty participants (twenty pairs; six men; 22.35 years old, range 18-28 years) are reported below. Theory of mind localizer data from two participants were excluded from the analysis due to excessive motion during this task.

### Materials

The materials were created in the same way as in Vanlangendonck et al. (2016). We manipulated the number, size and visibility of the relevant objects to create 6 conditions (Figure 3.1). In the privileged ground conditions (left column Figure 3.1), a competitor object was placed in a slot that was open only on the speaker's side. In communicative trials, speakers had to ignore this competitor object in order to unambiguously describe the target object, because speakers knew that the listener could not see the competitor object. In non-communicative blocks, speakers did not have to ignore the additional competitor object, because there was no listener present (see the procedure section below). We created two types of control conditions, in which there was no occluded competitor object. In the linguistic control conditions (middle column Figure 3.1), the occluded object was replaced by another, unrelated object. As a result, speakers saw one relevant object fewer in these conditions than in the privileged ground conditions. This condition is called the linguistic control condition, since speakers were expected to produce the same verbal response in this condition as in communicative privileged ground trials in which they successfully adjusted their response based on their addressee's perspective. In the visual control conditions (right column Figure 3.1), the object that was occluded in the privileged ground condition was visible to both participants. In these conditions, speakers therefore saw the same number of relevant objects as in the privileged ground conditions, hence we call these the "visual control" conditions.



**Figure 3.1:** Overview of a triplet of trials from the speaker's point of view, and the expected speaker responses in each of the six conditions. The task of the speaker was to describe a target object (red circle) for the listener (communicative blocks) or for him/herself (non-communicative blocks). In the communicative privileged ground condition, we expected speakers to take their addressee's perspective into account ("small glass"). If they did failed to take their addressee's perspective into account, they could also describe the target object from their own perspective ("medium glass"). There was no relevant perspective difference in the other conditions. In the linguistic control conditions, speakers were expected to give the same verbal response as in the communicative privileged ground condition. In the visual control conditions, both participants could see the competitor object that was occluded in the communicative privileged ground condition. Speakers thus saw the same number of relevant objects as in the privileged ground conditions. The green squares were added to the figure for clarification purposes to indicate the objects that differ between the privileged ground and the control conditions. They were not visible to the participants.

Twelve different empty virtual arrays were used in the experiment. Each array was filled with six to eight objects chosen from a total of 22 objects. Objects were selected from the Object Databank, stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, <http://www.tarrlab.org/>. Each object could appear in four different sizes to make sure that participants could not rely on absolute size. Target/competitor objects and filler objects all appeared in sets of one, two or three objects of the same type to make sure that participants could not predict which objects would be relevant. The speaker and listener always saw the same total number of objects in a trial because we added filler objects to the occluded slots if needed. We created a unique stimulus list for each participant pair. Trials were presented in blocks of six trials that were created using the same array. The order of the trials within each block was randomized, and we randomized the blocks of trials so that neighboring blocks did not use the same array. More information about the trial creation and randomization can be found in Vanlangendonck et al. (2016).

## Procedure

Participants were randomly assigned to the roles of speaker and listener by means of a coin toss. The speaker performed the task in the MRI scanner, while the listener was seated in

front of a computer in the MRI control room. The speakers spoke through a noise-cancelling microphone and listener could hear the speaker over headphones. Participants completed four blocks of the main task (2 communicative and 2 non-communicative blocks), followed by a Stroop task localizer<sup>1</sup> and a theory of mind localizer. The order of the four task blocks and the order of the localizers were counterbalanced. Each task block of the main experiment consisted of 60 trials, resulting in 240 trials in total. Before each block, speakers were informed whether the following block would be communicative or non-communicative.

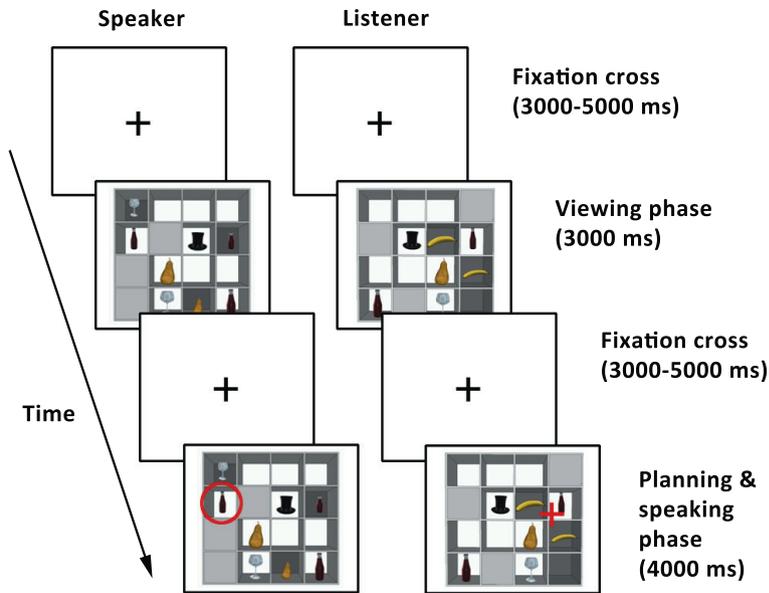
In communicative blocks, speakers and listeners played a referential communication game, in which the speaker described objects for the listener to select. Each trial featured a 4 x 4 array containing objects of different sizes. Each array contained three closed slots on each player's side, allowing us to manipulate which objects were in common ground. Objects that were visible to both players were in (visual) common ground, while objects that were only visible to one player were in that player's privileged ground (Figure 3.2). Each trial consisted of a viewing phase and a speaking phase. During the viewing phase, each player was shown his or her side of the array for 3000 ms. During the speaking phase (4000 ms), the speaker named the target object for the addressee (Figure 3.2). The target object was indicated using a red circle and was always in common ground. On the basis of the speaker's description, the listener clicked on the intended object in his or her display. A variable jitter of 3000-5000 ms preceded each phase of the trial.

The listener did not participate in the non-communicative blocks of the task. In these blocks, the speaker saw the same type of arrays as in the communicative blocks, but they were told that there was no listener and they were instructed to describe the cued object for themselves. Based on this instruction, we expected speakers to produce responses that did not take the distinction between common and privileged information into account in these blocks (Figure 3.1).

Speakers were instructed not to use spatial descriptions, such as 'leftmost' or 'third row'. Their verbal responses were recorded using a noise-cancelling microphone and they pressed a button when they were ready to start speaking, giving us a measure of the time they needed to plan their utterances. We recorded the location of the listeners' mouse clicks, which allowed us to calculate the listeners' accuracy. Before the start of the experiment, the speaker and the listener practiced the communicative task together using a real array and real objects. They then practiced the task together on a computer, and the speaker practiced the descriptive task by him/herself. The speaker also named all object pictures once before the start of the experiment to make sure they could easily recognize and name the objects.

---

<sup>1</sup>The Stroop task localizer produced null results so we did not include the results here.



**Figure 3.2:** Trial sequence of a communicative trial from the speaker and the listener's point of view. In the first phase of each trial, speakers and listeners each viewed their side of the array. In the second phase of the trial, a red circle indicated which object the speaker had to describe. The speaker planned their response, pressed a button once they were ready to start speaking and then described the target object for the listener. At the same time, the listener tried to click on the intended object. The trial sequence in the non-communicative blocks looked identical, except that the listener did not take part in this task.

After the experimental task, speakers completed a theory of mind localizer task (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2010; Saxe & Kanwisher, 2003; Van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). During this task, speakers were presented with twenty stories that required participants to represent false content. In half of the stories the false content concerned the physical state of an object (false photograph); in the other half of the stories it concerned another person's belief (false belief). Each story was presented for ten seconds, after which participants were given a statement about the story to judge. They had to respond to the statements with a button press within five seconds. A variable inter-trial interval of 4000-8000 ms preceded each trial.

### Data acquisition and analysis

Participants were scanned in a Siemens 3T Skyra scanner using a 32-channel head coil. The functional images were acquired using an EPI multi-echo sequence (TR = 2250 ms; TE1 at 9 ms, TE2 at 19.3 ms, TE3 at 30 ms, TE4 at 40 ms; 36 slices; ascending slice order; slice thickness = 3 mm; slice gap = 0.3 mm; 64 x 64 matrix size; field of view = 212 x 212 mm; flip angle = 90°; voxel size = 3.3 x 3.3 x 3 mm). A high-resolution T1 image was acquired using an MPRAGE

sequence (TR = 2300 ms; TE = 3.03 ms; 192 slices; voxel size = 1 x 1 x 1 mm, field of view = 256 x 256 x 192 mm).

Each of the four task blocks and the localizer tasks were scanned in separate runs. We acquired 30 additional functional scans before each block. These scans were used to calculate the optimal weighting of the five echoes, and this weighting matrix was applied to the remaining functional scans (Poser, Versluis, Hoogduin, & Norris, 2006). The functional images were processed using SPM8 (Statistical Parametric Mapping, [www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). The preprocessing of the functional images consisted of realignment to correct for head motion, slice timing correction to the onset of the middle slice, coregistration of the functional images to the T1 based on the subject-mean functional image, normalization to MNI space and spatial smoothing using a 3-dimensional isotropic Gaussian smoothing kernel (full-width half-maximum = 8 mm).

In the first-level statistical model, we included six event-types modeling the viewing phases (1 per condition), six event-types modeling the planning phases (1 per condition) and two event-types modeling the speaking phases (communicative and non-communicative). Events were modeled from picture onset until the button press for planning, and from the button press until picture offset for speaking. We also included six motion regressors per run. A separate statistical model was created to analyze the theory of mind localizer. This model included four event-types (false belief stories, false photograph stories, false belief statements and false photograph statements) and six motion regressors. We used boxcar functions to model the durations (10 s for stories; 5 s for statements). All event-types from the main experiment and the localizer were convolved with the hemodynamic response function. Individual t-contrasts were created and used in second-level random-effect analyses. Group analyses were performed using one-sample t-tests. Whole-brain results were corrected for multiple comparisons by combining a  $p < 0.001$  voxel-level threshold with a cluster extent threshold determined by means of a Monte Carlo simulation with 2500 iterations, after estimation of the smoothness of the data (Slotnick, Moo, Segal, & Hart, 2003). This revealed that clusters of 41 voxels or larger indicated statistically significant effects at the  $p < 0.05$  level, corrected for multiple comparisons.

For illustration purposes, we computed parameter estimates for regions of interest (ROIs) in the mPFC, and left and right TPJ using Marsbar (Brett, Anton, Valabregue, & Poline, 2002). The ROIs were based on the mPFC and TPJ clusters found in the *communicative privileged ground > communicative linguistic control* and the *communicative privileged ground > communicative visual control* contrasts. We selected only voxels that were significantly activated in both contrasts. The temporoparietal clusters were part of a large, interconnected cluster, so we further limited these ROIs to voxels with x coordinates under -35 (left) or over 35 (right).

We coded the sound files for adjective use and we used speakers' button press responses to determine the planning duration for each trial. We removed trials without a button press, trials without a response that could be coded and planning durations that were more than 3 standard deviations removed from the mean per subject from the dataset for the behavioral analysis (total 3.9 % of trials removed). Listener performance was calculated by determining whether listeners clicked on the right slot within the 4000 ms response interval. We then analyzed speakers' modifier use and planning durations using 2 x 3 repeated measures ANOVAs to investigate the effects of block (communicative or non-communicative), condition (privileged ground, linguistic control or visual control), and the interactions between these two factors. In addition, we tested whether we could replicate the results of our previous eye-tracking study (Vanlangendonck et al., 2016) by comparing the communicative privileged ground condition to each of the communicative control conditions. All behavioral analyses were run in R version 3.0.3.

## Results

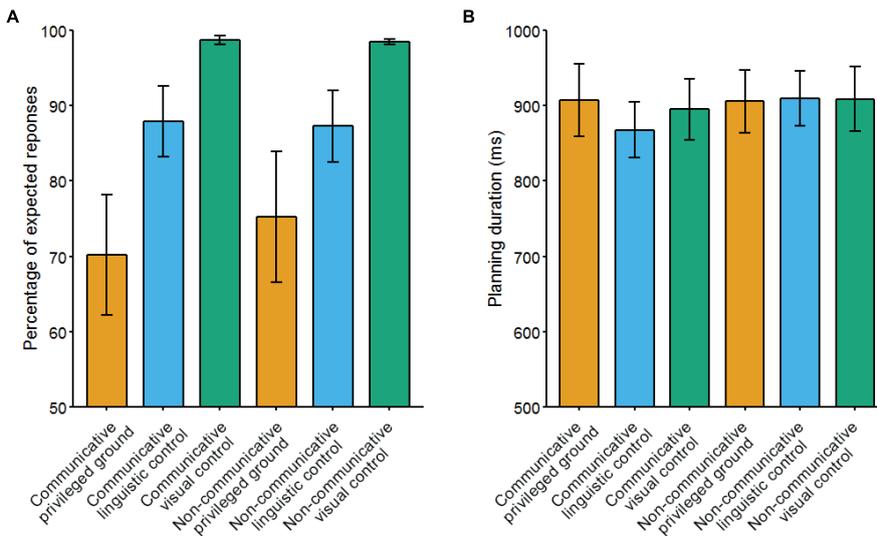
### Behavioral results

**Modifier use.** Figure 3.3A shows the mean percentage of "expected responses", i.e. responses that match the predictions in Figure 3.1. In the communicative privileged ground condition, we expected speakers to take their addressee's perspective into account, while we did not expect speakers to take their addressee's perspective into account in the non-communicative privileged ground condition. A 2 x 3 repeated-measures ANOVA with type of block (communicative or non-communicative) and condition (privileged ground, linguistic control or visual control) as factors revealed a significant main effect of condition ( $F(2,38) = 14.86, p < 0.001$ ), but no significant main effect of block ( $F(1,19) = 0.095, p = 0.76$ ) and no significant block x condition interaction effect ( $F(2,38) = 0.16, p = 0.71$  after Greenhouse-Geisser correction). Post-hoc paired t-tests revealed significantly more expected responses in the linguistic control conditions compared to the privileged ground conditions ( $t(19) = -2.97, p < 0.05$ ) and in the visual control conditions compared to the linguistic control conditions ( $t(19) = -2.52, p < 0.05$ ) after Holm-Bonferroni correction for multiple comparisons.

Speakers produced more expected responses in the communicative linguistic control condition compared to the communicative privileged ground condition,  $t(19) = 3.07, p < 0.01$ , and in the communicative visual control condition compared to the communicative privileged ground condition,  $t(19) = 3.68, p < 0.01$ . These findings are in line with previous studies with similar designs that showed that speakers generally take their addressee's perspective into account in communicative tasks, although they cannot completely ignore privileged ground information (Vanlangendonck et al., 2016; Yoon, Koh, & Brown-Schmidt, 2012; Wardlow Lane, Groisman, & Ferreira, 2006; Wardlow Lane & Ferreira, 2008).

**Planning duration.** An overview of the mean planning durations per condition can be found in Figure 3.3B. A 2 x 3 repeated-measures ANOVA with type of block (communicative or non-communicative) and condition (privileged ground, linguistic control or visual control) as factors did not reveal significant main or interaction effects (block:  $F(1,19) = 1.84$ ,  $p = 0.19$ ; condition:  $F(2,38) = 1.65$ ,  $p = 0.21$ ; block x condition:  $F(2,38) = 2.18$ ,  $p = 0.13$ ).

In order to compare these results with our previous eye-tracking study, we compared planning durations in the communicative privileged ground condition to each of the communicative control conditions using paired t-tests. We found that speakers took longer to plan their utterance in the communicative privileged ground condition compared to the linguistic control condition,  $t(19) = 2.53$ ,  $p < 0.05$ . We did not find a significant difference in planning duration between the communicative privileged ground condition and the communicative visual control condition,  $t(19) = 0.82$ ,  $p = 0.42$ .



**Figure 3.3:** Percentage of expected responses and mean planning durations per condition. The expected responses were coded based on the predictions in Figure 3.1. Planning durations were calculated from picture onset until speakers pressed the button to indicate that they were ready to respond. Error bars indicate standard error of the mean.

**Manipulation checks.** To check whether our block manipulation worked, we checked whether speakers indeed changed their strategy between the communicative and the non-communicative blocks. In the communicative blocks, we expected them to take their addressee's perspective into account in the privileged ground trials. Indeed, we found that speakers produced mostly responses that took their addressee's perspective into account in

the communicative privileged ground condition (70.17%). Given that there was no addressee present in the non-communicative blocks, speakers did not have to adapt their responses to what someone else could see in the non-communicative privileged ground conditions. In line with this prediction, we found that speakers produced descriptions from their own perspective in 75.22% of trials in the non-communicative privileged ground condition. We hence conclude that the communicative manipulation was successful.

In the communicative blocks, listeners tried to select the object described by the speaker. Overall listener performance was high (83.47% correct), indicating that speakers and listeners understood the task.

### **fMRI results**

The goal of the present study was to investigate the neural mechanisms underlying speakers' ability to take into account common ground during language production. First we compared language planning and viewing in a communicative and a non-communicative context. We hypothesized that speakers would use a mental model of their addressee when planning an utterance in the communicative as compared to the non-communicative blocks. In addition, given that keeping track of the distinction between objects in common and in privileged ground only matters in the communicative blocks, we expected that speakers might use different strategies to explore the visual arrays before knowing which object to describe depending on whether they were in a communicative or a non-communicative context. Second, and most importantly, we compared language production in situations in which the speaker has to take into account common ground in order to communicate efficiently with their addressee and situations in which this is not necessary. This we explored in the comparison between communicative versus non-communicative privileged ground planning, as well as by comparing the communicative privileged ground planning to the two control conditions (linguistic, visual) in the communicative blocks. For the "viewing" analysis, we focused on the viewing phase of the trials (see Figure 3.2). For all other analyses, we focused on the planning phase, i.e. the time between the moment the speaker saw the array with a red circle around the target object and the moment they pressed the button to indicate that they were ready to start speaking.

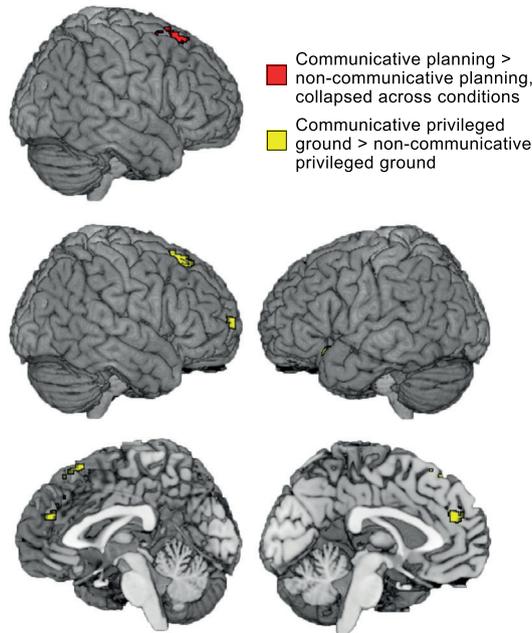
**Communicative vs. non-communicative privileged ground planning.** We first compared the communicative and the non-communicative privileged ground trials because speakers only had to take their addressee's perspective into account in the communicative privileged ground trials. The effect of communicative context may therefore be strongest when comparing these conditions directly. The contrast between planning in the communicative and non-communicative privileged ground conditions resulted in a series of clusters described in Table 3.1 and visualized in Figure 3.4. We found two right superior frontal clusters, one of which extends into the medial frontal gyrus, a left superior medial frontal

cluster that extends into the left superior frontal gyrus, a cluster in the left insula and inferior frontal gyrus and a right inferior frontal cluster that extends into the right insula.

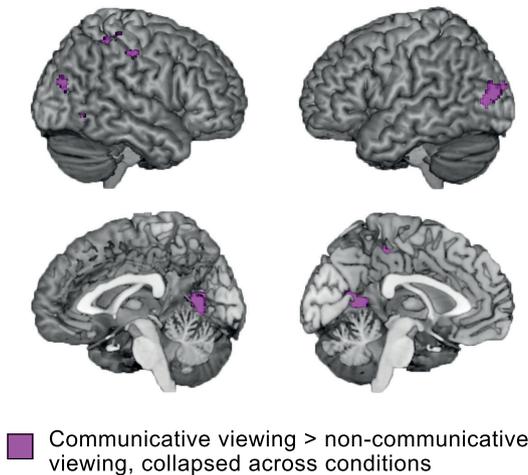
**Communicative privileged ground vs. linguistic and visual control planning.** We compared planning in the communicative privileged ground condition to each of the control conditions. The results of these contrasts overlap considerably, as can be seen in Figure 3.6. The contrast between the communicative privileged ground condition and the communicative linguistic control condition resulted in a large cluster covering parts of the inferior and superior parietal lobule as well as the superior occipital gyrus. In the frontal lobe, we found a superior medial frontal cluster, a right orbitofrontal cluster and left and right middle frontal clusters. Finally, we found bilateral insula and fusiform activations. Similarly, the contrast between the communicative privileged ground condition and the communicative visual control condition resulted in a large bilateral cluster covering the superior parietal lobule and angular gyri. In addition, we found a number of left and right inferior frontal, middle frontal and superior frontal activations. We also found clusters in the right fusiform gyrus and the right middle temporal gyrus, as well as in the left insula and the left cerebellum.

**Communicative vs. non-communicative planning.** We compared speech planning in the communicative and the non-communicative blocks, collapsed over conditions. This resulted in one cluster in the right superior frontal gyrus (Figure 3.4, Table 3.1).

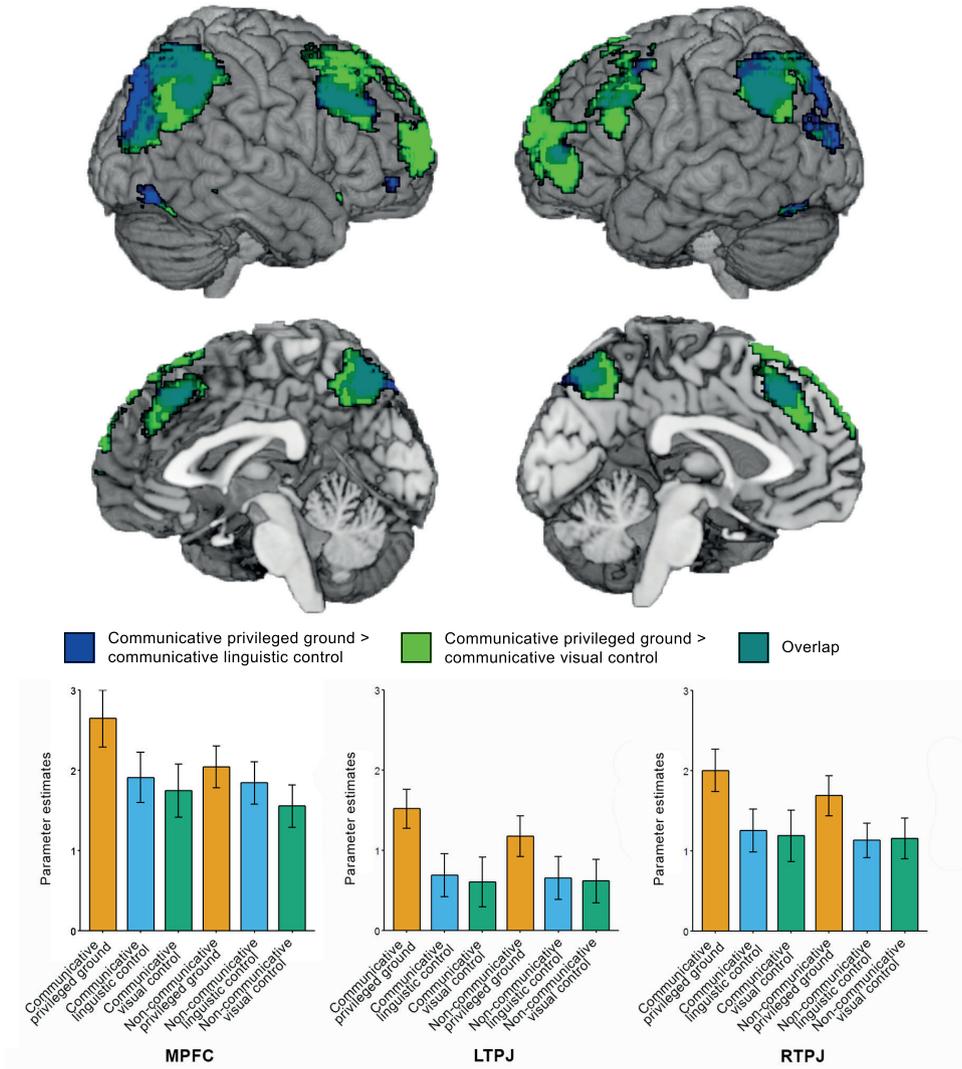
**Communicative vs. non-communicative viewing.** Given that common ground is thought to be updated continuously during conversation (Clark & Krych, 2004), we expected that we may find neural differences even before speakers start planning their utterances when comparing communicative and non-communicative task blocks. This comparison revealed clusters in the left and right middle occipital and calcarine gyri, as well as the right postcentral gyrus, the left cingulate cortex and the right inferior temporal gyrus (Figure 3.5, Table 3.1).



**Figure 3.4:** Brain areas showing greater activity in the communicative planning conditions compared to the non-communicative planning conditions (red) and the brain regions showing greater activity during the planning phase of the communicative privileged ground condition compared to the non-communicative privileged ground condition (yellow). In the latter comparison, the effect of communicative context was expected to be maximal.



**Figure 3.5:** Brain areas showing greater activity in the communicative viewing conditions compared to the non-communicative viewing conditions.



**Figure 3.6:** Brain areas showing greater activity during planning in the communicative privileged ground condition compared to the communicative linguistic planning condition (blue) and the communicative visual control condition (green). The graphs below depict the parameter estimates in the different conditions in the mPFC, and the left and right TPJ. Error bars represent standard errors of the mean. Note that we did not test for statistically significant differences between the parameter estimates of different conditions in order to avoid making the non-independence error.

**Table 3.1:** Whole-brain results for task contrasts of interest.

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>Communicative privileged ground planning &gt; non-communicative privileged ground planning</b>					
right superior frontal gyrus	219	6.37	16	22	60
right superior frontal gyrus		5.19	18	30	54
right superior medial frontal gyrus		4.51	8	24	60
left superior medial frontal gyrus	199	5.04	-8	46	24
left superior frontal gyrus		4.66	-12	26	40
right anterior cingulate cortex		4.50	6	44	24
left insula	139	4.83	-34	20	-8
left inferior frontal gyrus (pars orbitalis)		4.72	-38	22	-16
left inferior frontal gyrus (pars orbitalis)		4.17	-40	32	-10
right superior frontal gyrus	41	4.40	22	66	6
right inferior frontal gyrus (pars orbitalis)	45	4.33	40	24	-10
right insula		4.24	32	20	-6
<b>Communicative privileged ground planning &gt; communicative linguistic control planning</b>					
left inferior parietal lobule	6339	7.53	-48	-50	48
right superior occipital gyrus		7.50	36	-78	44
right superior parietal lobule		6.66	38	-58	60
left superior medial frontal gyrus	656	6.35	-8	30	42
right superior medial frontal gyrus		4.84	8	28	42
left supplementary motor area		4.38	-6	22	50
right middle frontal gyrus	619	5.92	48	30	34
right inferior frontal gyrus (pars opercularis)		5.10	50	20	38
right middle frontal gyrus		5.02	44	38	32
left middle frontal gyrus	394	4.93	-34	6	52
left middle frontal gyrus		4.71	-46	28	34
left middle frontal gyrus		4.67	-36	12	34
right insula	62	4.73	34	24	-6
right middle orbital gyrus	56	4.61	38	48	-8
right middle orbital gyrus		4.01	46	50	-8
right fusiform gyrus	162	4.55	38	-74	-18
right fusiform gyrus		3.95	42	-56	-16

**Table 3.1:** Whole-brain results for task contrasts of interest. (*continued*)

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
right lingual gyrus		3.87	32	-82	-18
left insula	97	4.48	-30	22	0
left middle frontal gyrus	94	4.47	-44	50	8
left middle frontal gyrus		3.68	-38	46	4
left fusiform gyrus	82	4.36	-38	-72	-18
<b>Communicative privileged ground planning &gt; communicative visual control planning</b>					
right angular gyrus	7803	7.15	54	-58	36
right superior parietal lobule		6.45	14	-64	58
right angular gyrus		6.43	36	-66	48
right middle frontal gyrus	3493	6.93	48	26	36
left superior middle gyrus		6.50	-4	32	38
right middle frontal gyrus		6.35	44	20	42
left middle frontal gyrus	1014	6.33	-38	12	36
left inferior frontal gyrus (pars triangularis)		5.03	-60	20	22
left middle frontal gyrus		5.01	-42	26	40
right superior frontal gyrus	653	6.12	34	62	14
right superior frontal gyrus		5.78	26	64	12
right middle frontal gyrus		4.93	36	64	2
right inferior frontal gyrus (pars orbitalis)	287	5.76	32	24	-8
right insula		4.68	32	28	2
right inferior frontal gyrus (pars orbitalis)		4.48	42	24	-16
right fusiform gyrus	133	4.68	40	-62	-20
right fusiform gyrus		4.05	34	-68	-16
right fusiform gyrus		3.73	26	-64	-12
right middle temporal gyrus	64	4.65	50	-42	-10
right middle temporal gyrus		4.30	58	-38	-10
right inferior temporal gyrus		3.88	58	-48	-10
left cerebellum	93	4.57	-36	-68	-22
left cerebellum		4.22	-30	-74	-22
left insula	65	4.33	-26	24	-4

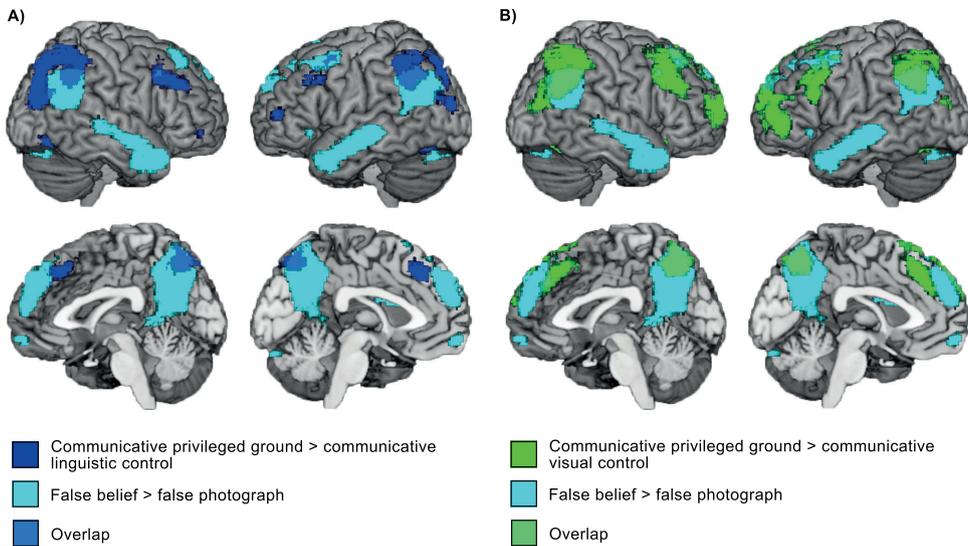
**Table 3.1:** Whole-brain results for task contrasts of interest. (*continued*)

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>Communicative all conditions planning &gt; non-communicative all conditions planning</b>					
right superior frontal gyrus	106	4.74	18	20	60
right superior frontal gyrus		4.30	20	6	64
right superior frontal gyrus		4.27	18	28	58
<b>Communicative all conditions viewing &gt; non-communicative all conditions viewing</b>					
left middle occipital gyrus	241	5.75	-42	-80	6
left middle occipital gyrus		4.67	-36	-86	12
left superior occipital gyrus		4.29	-20	-96	18
right postcentral gyrus	362	5.60	42	-28	42
right inferior parietal lobule		4.35	44	-42	54
right precentral gyrus		4.30	38	-20	44
left calcarine gyrus	224	5.16	-14	-64	14
left calcarine gyrus		5.10	-12	-58	8
left middle cingulate cortex	61	5.03	-8	-34	50
right calcarine gyrus	205	4.89	6	-64	10
right calcarine gyrus		3.76	10	-56	12
right inferior temporal gyrus	66	4.51	48	-60	-6
right inferior occipital gyrus		4.00	40	-68	-6
right middle occipital gyrus	65	4.22	40	-78	22

**Theory of mind localizer: false belief vs. false photograph.** We compared the false belief and false photograph statement conditions from the theory of mind localizer. In this contrast, we found a large set of brain regions commonly found in theory of mind tasks including the bilateral TPJ, the precuneus and the mPFC (Table 3.2, Figure 3.7).

**Table 3.2:** Whole-brain results for the comparison between false belief and false photograph statements.

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>False belief &gt; false photograph statements</b>					
left angular gyrus	1995	10	-56	-66	24
left middle temporal gyrus		7.77	-46	-56	22
left supramarginal gyrus		6.16	-60	-52	36
left precuneus	4297	9.7	2	-64	38
left precuneus		8.24	0	-56	40
left precuneus		7.53	-12	-50	40
right superior frontal gyrus	4999	8.69	16	46	34
right superior medial frontal gyrus		7.51	10	50	30
left superior frontal gyrus		7.2	-18	24	46
right angular gyrus	2143	8.38	48	-48	28
right angular gyrus		8.32	52	-62	26
right middle temporal gyrus		7.65	56	-60	18
left temporal pole	1453	7.23	-54	10	-32
left middle temporal gyrus		7.06	-62	-22	-12
left middle temporal gyrus		6.42	-54	-4	-20
right middle temporal gyrus	1565	6.95	54	4	-32
right temporal pole		6.59	50	20	-30
right middle temporal gyrus		6.56	58	-30	-2
right cerebellum	224	6.75	24	-80	-26
right cerebellum		4.5	44	-74	-24
left middle orbital gyrus	162	5.93	0	60	-12
left superior orbital gyrus		4.17	-14	58	-10
left middle orbital gyrus		3.81	-22	54	-10
left caudate nucleus	76	5.71	-14	8	20
left caudate nucleus		4.88	-12	-2	20
left cerebellum	395	5.67	-28	-76	-30
left cerebellum		5.67	-18	-88	-26
left cerebellum		4.87	-46	-72	-26
left inferior frontal gyrus (pars orbitalis)	46	4.81	-42	24	-8
left inferior frontal gyrus (pars orbitalis)		4.21	-48	28	-4



**Figure 3.7:** Brain areas showing greater activity during planning in the communicative privileged ground condition compared to the communicative linguistic control condition (A, blue), in the communicative privileged ground condition compared to the communicative visual control condition (B, green), and in the false belief condition compared to the false photograph condition in the theory of mind localizer (A & B, cyan).

## Discussion

In this study, we examined the neural mechanisms that allow speakers to adapt their language use to moment-to-moment changes in common ground by manipulating which information they do or do not share with their addressee, in addition to comparing language production in a communicative and a non-communicative context.

### Behavioral results

Speakers generally took their addressee's perspective into account when designing referring expressions in the communicative blocks, although they failed to ignore privileged information on all trials. These results are in line with previous findings using similar paradigms (Vanlangendonck et al., 2016; Wardlow Lane et al., 2006; Wardlow & Ferreira, 2008; Yoon et al., 2013) and mostly results replicate the findings of our previous eye-tracking study (Vanlangendonck et al., 2016). Speakers produced slightly less expected responses in the communicative conditions in the current study compared to the previous study. One likely cause for this difference is that speakers switched between communicative and non-communicative blocks in the current experiment, which may have made the task more difficult. Another difference is that in the current study, speakers' planning durations were shorter in the communicative linguistic control condition than in the communicative

privileged ground condition. Comparing the planning durations across conditions (Figure 3.3B) suggests that speakers were especially fast in this condition compared to the other conditions.

### **fMRI results: communicative privileged ground vs. non-communicative privileged ground planning**

Speakers were only expected to adjust their language use based on their addressee's perspective during speech planning in the communicative privileged ground condition. We therefore expected that the effect of communicative context on speech planning may be strongest in the direct comparison between speech planning in the communicative and the non-communicative privileged ground conditions. This contrast revealed clusters of activation in the right superior frontal gyrus, the left insula and inferior frontal gyrus, and in the dorsal mPFC.

The mPFC is a core region of the mentalizing network that may support the integration of social information over time (Van Overwalle, 2009). It is thought to be subdivided into a more dorsal section, which is activated when thinking about the mental states of dissimilar others, and a ventral part, which is activated especially when thinking about the mental states of similar others (Mitchell, Macrae, & Banaji, 2006; Van Overwalle, 2009). Given that speakers and listeners did not know each other before the start of the experiment and that successful performance in the communicative privileged ground condition required speakers to focus on differences in perspective, it is unsurprising that the dorsal part of the mPFC was activated in this comparison. Our whole-brain results suggest that the mPFC is selectively engaged by the communicative privileged ground condition. The parameter estimates plotted in Figure 3.6 illustrate this. The mPFC thus appears to be especially activated when speakers perform a communicative task that requires them to adapt their language use to their addressee's needs in order to communicate efficiently. These findings mimic the results of a comprehension study that found that adult listeners only show increased activation in the dorsal mPFC when they have to take a speaker's perspective into account to respond appropriately (Dumontheil, Hillebrandt, Apperly, & Blakemore, 2012).

### **fMRI results: communicative privileged ground vs. linguistic and visual control planning**

When comparing the communicative privileged ground condition to each of the control conditions, we found a large, overlapping network of activations that include the core regions of the mentalizing network (Van Overwalle, 2009; Schurz et al., 2014): the mPFC and bilateral TPJ.

We found activations bilaterally in the dorsal/posterior part of the TPJ in this contrast. The TPJ has been proposed to be important for inferring temporary states of other people, such as goals intentions and desires, even when they differ from your own (Van Overwalle, 2009). The posterior/dorsal part of this area may be especially important for the processing of mental perspectives (Schurz et al., 2014). Speakers' visual perspective for relevant objects only differed from their addressee's in the communicative privileged ground condition. The cluster we find in this area may therefore be the result of representing this perspective difference.

In addition to regions involved in mentalizing, we found bilateral clusters of activation in the ventro- and dorsolateral prefrontal cortex, which are known to play an important role in cognitive control processes (e.g., Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004; Petrides, 2005). In the communicative privileged ground condition, speakers had to ignore an occluded competitor object. In contrast, all relevant objects were mutually visible in the control conditions. Previous studies (Brown-Schmidt, 2009; Nilsen & Graham, 2009; Wardlow, 2013) have shown that inhibitory control skills correlate with people's ability to take another person's perspective into account during social interaction, suggesting that adapting your language use to another person's perspective depends on your ability to inhibit your own perspective.

### **fMRI results: communicative vs. non-communicative planning**

When comparing brain activity during speech planning in the communicative and non-communicative blocks, we found activation in a right superior frontal cluster. Although not considered one of the core components of the mentalizing network, it is interesting to note that this cluster appears in all contrasts we tested, as well in the theory of mind localizer. Similar right superior frontal activations have been found before in theory of mind tasks, including false belief tasks and tasks in which participants made trait judgments (Schurz et al., 2014).

### **fMRI results: communicative vs. non-communicative viewing**

The contrast between communicative and non-communicative viewing revealed a set of occipital and parietal clusters. One likely explanation for these findings is that speakers used a different strategy in the communicative and non-communicative viewing phases. The viewing phase allowed speakers to visually explore the objects in the array before seeing the cue indicating which object they had to describe. However, the distinction between objects in common and in privileged ground only mattered in the communicative blocks. Speakers may have therefore paid additional attention to the location of objects (in open or occluded slots) in the communicative blocks, resulting in increased activity in areas associated with visual attention (Corbetta & Schulman, 2002).

### **fMRI results: theory of mind localizer**

It is interesting to note that the clusters of activation we found in the comparisons between the conditions in the main task only partially overlap with the results of the theory of mind localizer. The temporoparietal clusters we found in the main task extend more dorsally compared to the cluster we found in the theory of mind localizer task, and we find little overlap between clusters in the mPFC. One possible explanation for these differences is that the theory of mind localizer we used here is not the most suitable task to tap into the perspective-taking processes speakers engaged in during the main task. Adapting your language use based on what your addressee can see requires relatively low-level perspective-taking (so-called level 1 perspective-taking). While both visual perspective-taking and false belief tasks require the representation of different perspectives, the neural correlates of these tasks do not completely overlap (Schurz, Aichhorn, Martin, & Perner, 2013). Alternatively, the limited overlap between the task contrasts and the localizer results may be due to statistical thresholding. In line with this possibility, we found increased overlap when we lowered the statistical threshold, especially in the TPJ.

### **General discussion**

In line with previous studies, our results show that adjusting your linguistic message for an addressee engages the mentalizing or theory of mind network. However, this study is the first to tease apart the effects of communicative context and the need to adjust your linguistic utterance to take common ground into account. Our results suggest that the mentalizing network plays a crucial role when speakers have to consider which information they share with their addressee in order to be informative. We extend previous findings by showing that the mPFC does not appear to be sensitive to communicative context per se, but rather becomes more activated when the communicative context has consequences for linguistic processing (i.e. when speakers have to take common ground into account to communicate efficiently). The TPJ, on the other hand, may be important for processing and representing your interlocutor's perspective when it differs from your own. It may therefore be especially sensitive to potentially relevant perspective differences.

What do these findings mean for accounts of the neurobiology of language? One important finding is that the mentalizing network appears to be especially involved during communicative language processing when speakers have to take common ground into account. Our findings thus do not suggest that the mentalizing network always comes online when speakers design utterances for an addressee. Rather, areas involved in social cognition appear to be selectively activated when speakers need to take common ground into account in order to communicate efficiently. However, our results from the viewing phase show that being in a communicative as compared to a non-communicative context

can lead to neural differences before language planning has started (i.e., before speakers knew which object they had to describe). While we did not collect eye-tracking data, a likely explanation is that speakers were more sensitive to the distinction between common and privileged information in the communicative viewing phases as compared to the non-communicative viewing phases. A recent MEG study (Stolk et al., 2013) also reported neural differences between communicative and non-communicative settings before participants were presented with a communication problem. Combined, our results suggest that speakers monitored which objects were visible to their addressee during the viewing phase in the communicative blocks and then used this information to determine whether they had to adjust their linguistic utterance to take their addressee into account. This suggests that speakers in a communicative setting continuously update common ground and use this information to adapt their linguistic utterances based on their addressee's needs.

In the current experiment, speakers could adapt to what their addressee was saying relatively easily by considering which objects were visible to the other person. However in most everyday settings, taking common ground into account also requires considering what your interlocutor knows and feels. Future research will have to clarify how the findings from the current study relate and extend to communicative tasks that require more high-level perspective taking (e.g., considering what your interlocutor knows about a topic). In addition, it is important to note that during real-life conversations, the distinction between communicative situations in which speakers have to take common ground into account and situations in which this is not necessary is much less clear-cut than in the current study.





Collaborative learning engages  
different memory retrieval networks



## Abstract

Learning often occurs in communicative and collaborative settings, yet almost all research into the neural basis of memory relies on participants encoding and retrieving information on their own. We investigated whether learning linguistic labels in a collaborative context at least partly relies on cognitively and neurally distinct representations, as compared to learning in an individual context. Healthy participants learned labels for sets of abstract shapes in three different tasks. They came up with labels with another person in a collaborative communication task (*collaborative condition*), by themselves (*individual condition*), or were given pre-determined unrelated labels to learn (*arbitrary condition*). Immediately after learning, participants retrieved and produced the labels aloud in the MRI scanner. The fMRI results show that the retrieval of collaboratively generated labels as compared to individually learned labels engages brain regions involved in understanding others (*mentalizing*). This study is the first to show that collaboration during encoding affects the neural networks involved in retrieval.

Based on:

Vanlangendonck, F., Takashima, A., Willems, R. M., & Hagoort, P. (submitted). Collaborative learning engages different memory retrieval networks.

## Introduction

Learning often occurs in communicative and collaborative settings, yet almost all research into the neural basis of memory relies on participants encoding and retrieving information on their own. Intuitively, it may seem only a small step to extrapolate that the memory systems that support encoding and retrieval in an individual context are also involved in encoding and retrieving in a collaborative context. However, there is tentative evidence from lesion research that suggests that learning in a collaborative context and learning in an individual context may at least partly rely on cognitively and neurally distinct processes.

Duff, Hengst, Tranel, & Cohen (2006) instructed patients with hippocampal amnesia to perform multiple rounds of the “tangram task” (Clark & Wilkes-Gibbs, 1986) with a communication partner. In this task, a director and a matcher each have a set of abstract figures, but their view of the other person’s figures is occluded by a barrier. The director instructs the matcher to rearrange their figures to match the director’s order. Over multiple rounds of this task, healthy participants converge on a set of labels that are used to refer to the abstract figures. These labels become shorter and more efficient with each repetition of the task (Clark & Wilkes-Gibbs, 1986). Interestingly, the amnesic patients could successfully learn novel labels for the figures over multiple rounds of the tangram task, but they were unable to learn arbitrary relations between labels and pictures in an individual paired-associate learning task. These findings suggest that learning labels in a collaborative context may rely on cognitively and neurally distinct representations compared to learning in a non-communicative context.

In the current study, we directly compared the neural representation of collaboratively generated linguistic labels to memory for labels that were encoded individually. During a behavioral session, healthy participants generated labels for abstract figures together with another person in a collaborative communication task (collaborative condition), by themselves (individual condition), or were given pre-determined arbitrary labels to learn by themselves (arbitrary condition). Immediately after learning, participants completed a communication task in the MRI scanner during which they retrieved the labels they had learned during the training session. This design allowed us to directly compare the retrieval of labels learned in the collaborative, individual and arbitrary contexts.

In line with the results in the hippocampal amnesia patients, we expected that the retrieval of pre-determined arbitrary labels should rely strongly on the hippocampus, while the retrieval of the collaboratively generated labels may be less hippocampus-dependent. More specifically, we hypothesized that the retrieval of labels learned in the collaborative condition would partly rely on brain areas involved in mentalizing, which are consistently found in studies of social cognition and interaction (Amodio & Frith, 2006; Frith & Frith, 2006;

Frith & Frith, 2010; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Van Overwalle, 2009). In addition, we expected that retrieving self-generated labels (collaborative and individual conditions) as compared to arbitrary labels may activate areas known to be important for storing semantic associations such as the left angular gyrus, middle temporal gyrus and temporal poles (Binder, Desai, Graves, & Conant, 2009). Behaviorally, we expected to replicate the results of previous tangram task studies during the behavioral practice phase. That is, speakers were expected to produce shorter and more efficient descriptions with each round of this task (Clark & Wilkes-Gibbs, 1986).

## **Method**

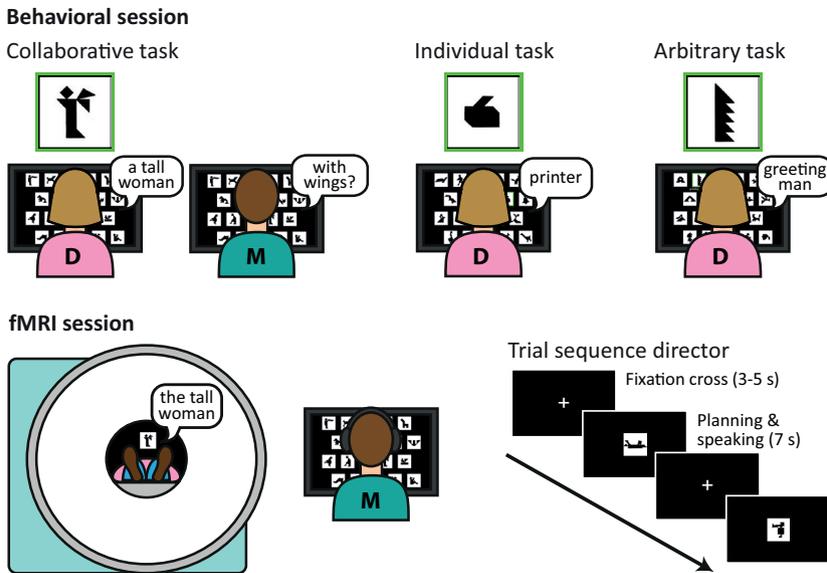
### **Participants**

36 pairs of native Dutch speakers participated in the experiment. All participants had normal or corrected-to-normal vision and no history of neurological disease. They gave written informed consent before the start of the experiment. Data from one pair were excluded due to excessive movement by the participant in the MRI scanner. The results of the remaining seventy participants (directors: 7 men, 21.43 years old, range 18-33 years; matchers: 6 men, 21.34 years old, range 18-31 years) are reported below.

### **Procedure**

The experiment consisted of two parts: a behavioral session and an fMRI session. The two parts were completed immediately after each other. Before the start of the experiment, participants were randomly assigned to the roles of director and matcher for the duration of the experiment by means of a coin toss. During the behavioral session, participants were seated behind different monitors in the same room. The monitors were separated by a screen, so the participants could not see each other or each other's monitor. The director learned sixty labels for pictures in three different tasks: a collaborative task, an individual task and an arbitrary task. During the fMRI session, the director performed a referential communication task with the matcher. We recorded the directors' vocal responses during both sessions. The order of the tasks and the matching between the tasks and the picture sets were counterbalanced. An overview of the tasks is depicted in Figure 4.1.

During the collaborative task, the director and matcher were seated at different computer monitors. Each participant saw the same set of twenty pictures on their monitor. On each trial, the director was cued to describe one of the pictures, and the matcher tried to click on the intended picture. The director and matcher could communicate freely during the entire task, and could discuss each picture as long as needed until the matcher had selected the right picture. Both participants received feedback on screen ("right" or "wrong") when the matcher clicked on a picture. Once the matcher had clicked on the described picture, the



**Figure 4.1:** Set-up of the experiment. The top row depicts the three tasks participants performed during the behavioral learning session. Note that the matcher only participated in the collaborative task. The described tangram pictures are shown at the top to illustrate the relationship between the pictures and the labels. After completing these three tasks, the director retrieved all previously learned labels in the MRI scanner, while the matcher tried to click on the described pictures. The trial sequence depicts the task in the MRI scanner from the director's point of view. Each picture was presented for seven seconds, during which the director planned and produced his/her response out loud. The director was instructed to press a button just before speaking.

director was cued to describe the next picture. Participants completed four rounds of the collaborative task. In each round of the task, the director was cued to describe each of the twenty pictures once. The locations of the pictures on screen as well as the order in which they were cued were randomized before each round. Afterwards, the director received a list containing all pictures and wrote down the labels they had come up with below the corresponding pictures.

The matcher did not participate in the individual task, but was seated in the same room. In this task, the director saw a different set of twenty pictures on their monitor. The director was instructed to come up with a label for each picture that he/she could easily remember. He/she moved from picture to picture by pressing a button. In each round of the task, the director was cued to describe each of the twenty pictures once out loud. Directors completed four rounds of this task. The locations of the pictures on screen as well as the order in which they were cued were randomized before each round. Afterwards, the director received a list containing all pictures and wrote down the labels he/she had come up with below the corresponding pictures.

In the arbitrary task, the director was again presented with a different set of twenty pictures on his/her monitor. The director was instructed to learn a predetermined label for each picture. The matcher did not participate in this task. The director could move from picture to picture by pressing a button. When he/she pressed the button once, a cue appeared around one of the pictures. When they pressed the button again, the label they had to learn appeared below the picture. In the first round of the task, they could look at the pictures and labels without describing the pictures. From the second round onwards, they were instructed to say the label they had learned for each picture out loud when the cue appeared. When they were sure they knew the label associated with a specific picture, they could remove the picture from the practice set by pressing a second button. The directors practiced the labels until they had removed all labels from the practice set. Then they were asked to label all pictures once more to check their performance.

The matcher had to study all the labels the director had learned during the behavioral session before the start of the fMRI session to allow participants to successfully perform the referential communication game together during the fMRI session. The matcher therefore learned the same sets of labels as the director during the behavioral session. In the collaborative task, the director and matcher established and learned the labels together over the course of the task. The matcher received the labels the director learned in the other tasks on paper and studied these in silence. He/she was given the list of pictures and corresponding labels the director wrote down after the individual task and a list of the pictures and corresponding arbitrary labels the director had studied.

During the fMRI session, the director lay in the MRI scanner, while the matcher was seated at a computer outside the scanner room. The director spoke through a noise-cancelling microphone and could press a button on a button box; the matcher wore headphones and used a mouse to click on pictures. Participants were told that they would play a communication game in the scanner in which the matcher had to select the pictures named by the director. On each trial, the director was presented with one of the sixty pictures he/she had studied during the behavioral session. The director was instructed to press a button once he/she remembered the associated label and to then say this label out loud. The button presses provided us with a measure of the director's planning duration per trial. At the same time, the matcher saw 20 pictures, organized per set, and tried to click on the described picture. Trials for the director and matcher always started simultaneously, but differed in length. The director's trials lasted 7 seconds and were followed by a jittered ITI of 3-5 seconds. The matcher's trials lasted 9.5 seconds and were followed by an ITI of 0.5-3 seconds to give them sufficient time to select the right picture. All pictures were shown twice during the MRI task in randomized order, resulting in 120 trials in total. Before the director went into the scanner, we explained to both participants that they had learned the same sets of labels.

## Materials

The materials consisted of sixty abstract tangram figures, which were adapted from Read (1965). We ran an online pretest in which 20 participants rated 83 tangram pictures. They were asked to come up with a label for each picture and rated how difficult it was to describe each picture on a scale from 1 to 5. Based on these ratings, we created three sets of twenty tangram pictures that were matched based on the difficulty ratings and the predominant semantic category of the responses (animal, human or object responses).

The labels directors studied in the arbitrary task were selected from the labels participants produced in the same pretest as well as in a previous pretest in which we used the same task but a different set of tangram pictures. We only included labels that were generated for pictures that were not selected for the final stimulus set. The three sets of arbitrary labels were matched in terms of semantic categories as well as the average number of words per label. Each arbitrary label was then linked to an unrelated tangram picture. An example of an arbitrary picture-label pair is shown in Figure 4.1.

## Data acquisition and analysis

Participants were scanned in a Siemens 1.5T Avanto MRI scanner using a 32-channel head coil. Functional images were acquired using a T2\*-weighted gradient multi-echo planar imaging sequence (TR = 2320 ms; TE1 at 9.4 ms, TE2 at 21 ms, TE3 at 33 ms, TE4 at 44 ms, TE5 at 56 ms; 37 slices; ascending slice order; 3 mm slice thickness; 0.51 mm slice gap; 64 x 64 matrix size; 212 x 212 mm FOV; 90° flip angle and 3.3 x 3.3 x 3 mm voxel size). In addition, T1-weighted anatomical scans with 1 mm isotropic resolution were acquired (TR = 2250 ms; TE = 2.95 ms; 15° flip angle; 256 x 256 x 176 mm FOV).

We acquired 35 pre-scans before the start of the task in the MRI scanner. These scans were used to calculate the optimal weighting to combine the five echoes to one value per volume for each voxel, and this weighting matrix was applied to the remaining functional scans (Poser, Versluis, Hoogduin, & Norris, 2006). Preprocessing was done in SPM8 (Statistical Parametric Mapping, [www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). The preprocessing of the functional images consisted of slice timing correction to the onset of the middle slice, coregistration of the functional images to the T1 based on the subject-mean functional image, normalization to MNI space (resulting voxel size 2 x 2 x 2 mm) and spatial smoothing using a 3-dimensional isotropic Gaussian smoothing kernel (full-width half-maximum = 8 mm).

We were interested in testing whether directors' responses became shorter with each round of the collaborative task during the behavioral practice session. We therefore calculated how long each trial of the collaborative task lasted based on the time between stimulus presentation and the correct click by the matcher. In addition, we transcribed and coded

all responses produced by the director during the fMRI task. The responses were coded as correct or incorrect by comparing them to the arbitrary labels and the labels the directors wrote down during the behavioral session. Furthermore, we used the button press responses in the scanner to estimate planning durations in the scanner and counted the number of words per response. Two participants forgot to press the button during the task in the MRI scanner and one participant pressed the button at the wrong time (after instead of before naming the picture labels). We excluded their MRI button press data from the behavioral analysis, but included their fMRI data. We manually calculated the planning durations for these three participants by inspecting the voice onset time in the audio recordings of their responses using Praat (Broersma, 2001), and used these values for the fMRI analysis. For the analysis of the number of words per label and the planning durations, we excluded incorrect trials, trials without a button press and trials with planning durations that were more than 3 SD from the mean per condition and the mean per participant. Greenhouse-Geisser correction was applied when the sphericity assumption was violated and Holm-Bonferroni correction for multiple comparisons was used for all reported post-hoc tests.

We included five variables in the first-level general linear models: collaborative retrieval, individual retrieval, arbitrary retrieval, speaking and trials of no interest. The retrieval regressors were stick functions time-locked to picture onset. We included the planning duration associated with each trial as a linear parametric modulator. The speaking regressor was modeled as a stick function and we included the number of words per trial as a linear parametric modulator. The condition of no interest regressor included trials without a button press and trials with responses that did not match the labels learned during the behavioral session. Regressors were convolved with the hemodynamic response function. In addition, we included 24 nuisance regressors: the six realignment parameters, their square, their first derivative, and the realignment parameters used to realign the previous volume. Individual t-contrasts were created and used in second-level random-effect analyses. Group analyses were performed using one-sample t-tests. Whole-brain results were corrected for multiple comparisons by combining a  $p < 0.001$  voxel-level threshold with a cluster extent threshold determined by means of a Monte Carlo simulation with 2500 iterations, after estimation of the smoothness of the data (Slotnick, Moo, Segal, & Hart, 2003). This revealed that clusters of 41 voxels or larger indicated statistically significant effects at the  $p < 0.05$  level, corrected for multiple comparisons.

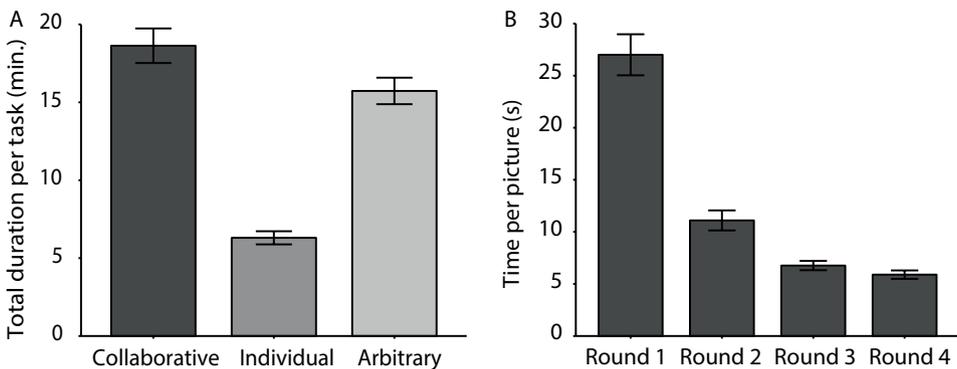
## Results

### Behavioral results

First, we tested whether the tasks during the behavioral session differed in duration. In addition, we tested whether we could replicate the results of previous studies that used the tangram task. These studies typically find that directors' descriptions become shorter and more efficient over time (e.g., Clark & Wilkes-Gibbs, 1986; Brennan & Clark, 1996). We therefore compared the time per picture across the four rounds of the collaborative training task. Second, we compared the planning durations, accuracy and number of words per label between conditions in the fMRI session.

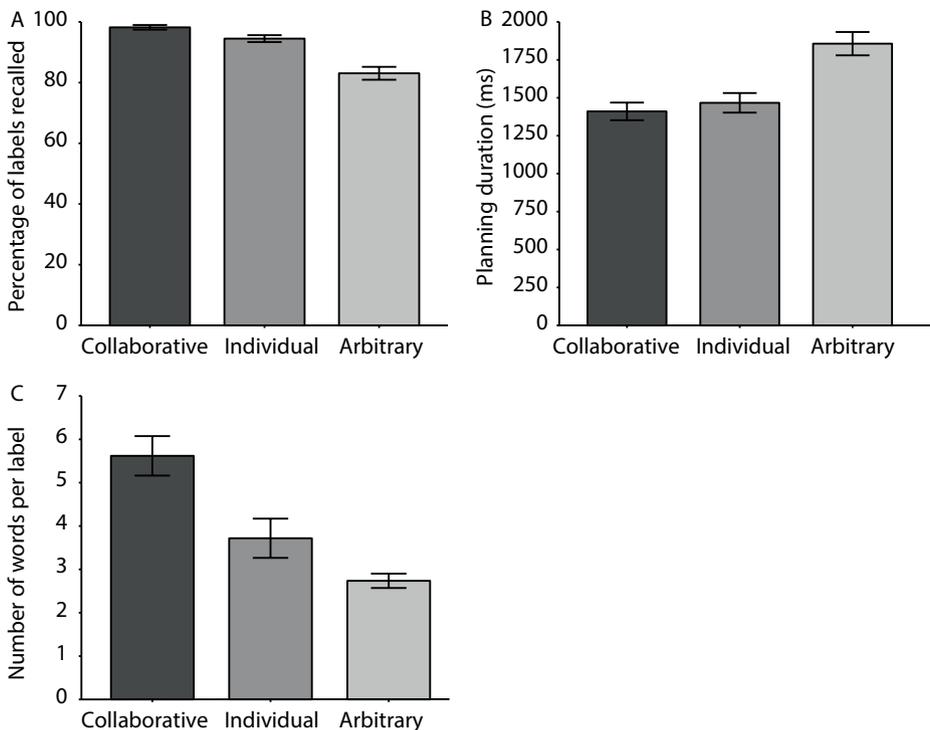
**Behavioral session.** The tasks during the behavioral session were all self-paced, and there were differences in duration between the tasks (see Figure 4.2A). A repeated-measures ANOVA with condition as within-subject factor revealed a significant main effect of condition (collaborative, individual or arbitrary) on the total duration per task,  $F(2,68) < 0.001$ . Post-hoc comparisons revealed significant differences between all conditions ( $p < 0.001$  for collaborative vs. individual and individual vs. arbitrary,  $p < 0.05$  for collaborative vs. arbitrary).

The duration per picture decreased with each round of the collaborative task (round 1: 27 s, round 2: 11.10 s, round 3: 6.76 s, round 4: 5.89 s; see Figure 4.2B). A repeated-measures ANOVA with round as within-subject factor revealed a main effect of round,  $F(3,102) = 120.58$ ,  $p < 0.001$ . A polynomial trend analysis revealed significant linear and quadratic trends across rounds (linear:  $F(1,34) = 145.08$ ,  $p < 0.001$ ; quadratic:  $F(1,34) = 92.85$ ,  $p < 0.001$ ). Our results thus replicate previous studies that found that directors' descriptions become more efficient and shorter with each round of the tangram task.



**Figure 4.2:** Behavioral results from the behavioral session: A) Total time per training task, and B) time per picture per round in the collaborative task. Error bars indicate standard error of the mean.

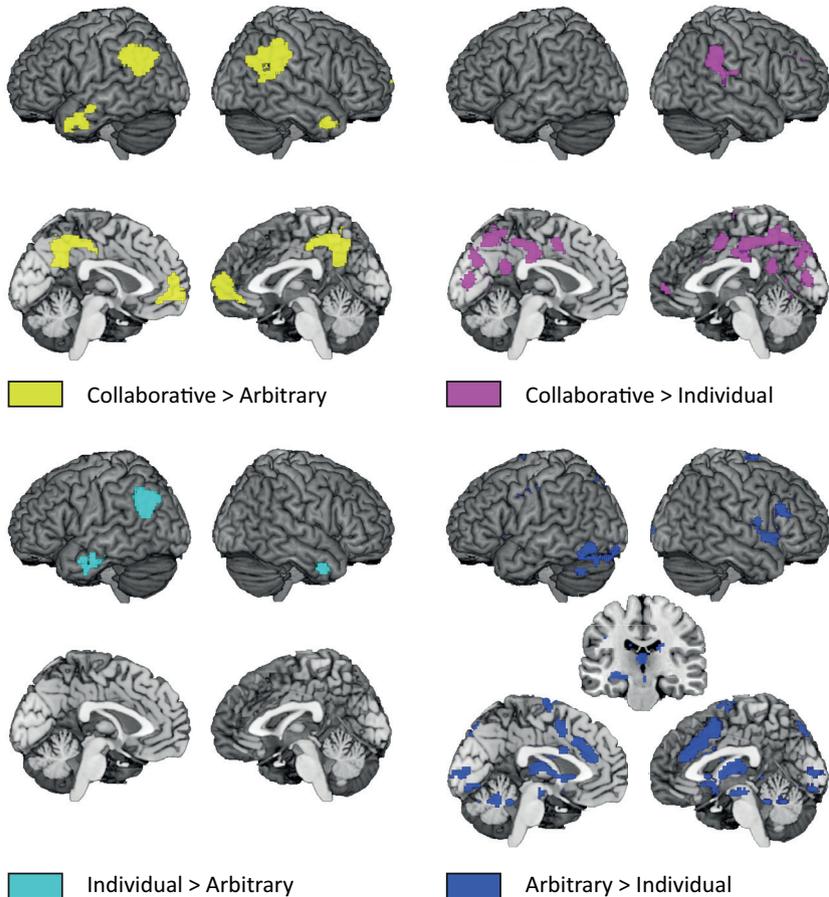
**fMRI session.** We performed repeated-measures ANOVAs with condition (collaborative, individual or arbitrary) as within-subject factor and accuracy, planning durations and number of words per label as dependent measures (see Figure 4.3). We found a significant main effect of condition on accuracy,  $F(2,68) = 34.42$ ,  $p < 0.001$ . Post-hoc comparisons revealed that accuracy was significantly higher in the collaborative condition compared to the individual condition ( $t(34) = 2.81$ ,  $p < 0.01$ ) and in the individual compared to the arbitrary condition ( $t(34) = 4.91$ ,  $p < 0.001$ ). We found a significant main effect of condition on planning durations in the MRI task,  $F(2, 68) = 72.26$ ,  $p < 0.001$ . Post-hoc comparisons revealed significantly shorter planning durations in the individual compared to the arbitrary condition,  $t(34) = 9.40$ ,  $p < 0.001$ . The difference between the collaborative and the individual conditions was not significant,  $t(34) = 1.55$ ,  $p = 0.13$ . We found a significant main effect of condition on the number of words per label in the MRI task,  $F(2, 68) = 26.08$ ,  $p < 0.001$ . Post-hoc comparisons revealed that responses contained significantly more words in the collaborative condition compared to the individual condition ( $t(34) = 4.61$ ,  $p < 0.001$ ) and in the individual compared to the arbitrary condition ( $t(34) = 2.70$ ,  $p < 0.05$ ).



**Figure 4.3:** Behavioral results from the fMRI session: A) percentage of labels recalled during the fMRI session, B) planning duration during the fMRI session (time from picture onset until the director's button press, and C) number of words per label. Error bars indicate standard error of the mean.

### fMRI results

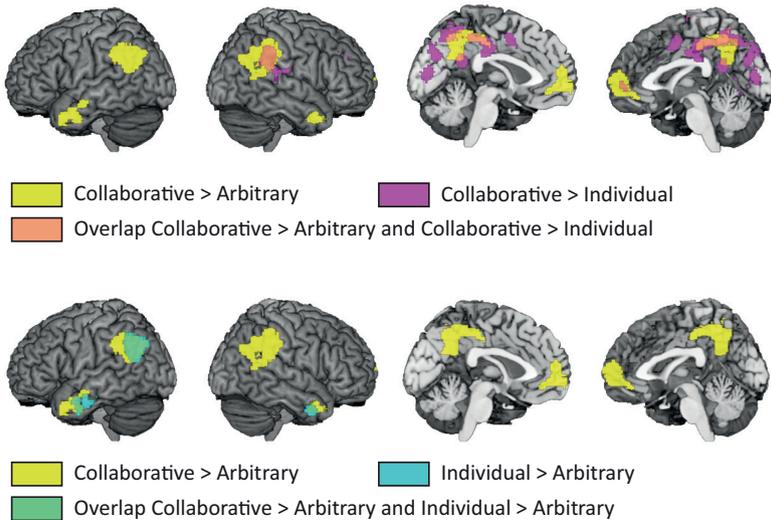
We compared the retrieval of labels learned in the collaborative condition to labels learned in the other conditions to examine the neural mechanisms of collaborative learning. In addition, we compared the retrieval of labels learned in the individual and arbitrary conditions to examine the effects of learning self-generated labels. The results of the fMRI analyses are listed in Table 4.1 and presented in Figures 4.4 and 4.5. The fMRI results are all from the retrieval phase of the trials between picture onset and the director's button press.



**Figure 4.4:** Brain regions showing increased activity during the retrieval of labels learned in different tasks by the director: the retrieval of collaboratively learned labels versus labels learned in the arbitrary task (yellow), the retrieval of collaboratively versus individually learned labels (violet), the retrieval of individually learned labels versus labels learned in the arbitrary task (light blue), and the retrieval of labels learned in the arbitrary task versus individually learned labels (dark blue).

**Collaborative > arbitrary.** The comparison between the collaborative and arbitrary conditions resulted in clusters in the precuneus, the bilateral angular gyri, medial prefrontal cortex and bilateral temporal poles (Figure 4.4, yellow).

**Individual > arbitrary.** The contrast between the individual and the arbitrary conditions revealed clusters in the left angular gyrus, and in the temporal poles extending into the middle and inferior temporal gyri (Figure 4.4, light blue).



**Figure 4.5:** Overlap between the task contrasts depicted in Figure 4.4. Orange indicates the overlap between the retrieval of collaboratively learned labels versus labels learned in the arbitrary task (yellow) and the retrieval of collaboratively versus individually learned labels (violet). Green indicates the overlap between the retrieval of collaboratively learned labels versus labels learned in the arbitrary task (yellow) the retrieval of individually learned labels versus labels learned in the arbitrary task (light blue).

**Collaborative > individual.** The contrast of main interest was between the collaborative and individual conditions. This comparison revealed a series of clusters including a large cluster connecting the midcingulate cortex and precuneus, a right temporoparietal cluster, a cluster in the left putamen, bilateral middle frontal clusters and several medial frontal clusters (Figure 4.4, violet).

**Arbitrary > individual.** The comparison between the arbitrary and individual conditions resulted in a large number of clusters including the anterior cingulate cortex, bilateral caudate nucleus, left hippocampus, bilateral insula, medial frontal cortex and the bilateral calcarine gyri (Figure 4.4, dark blue).

**Table 4.1:** Whole-brain results for the comparisons between the collaborative, individual and arbitrary conditions.

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>Collaborative &gt; Arbitrary</b>					
precuneus	2554	6.41	0	-46	48
right precuneus		6.12	8	-52	30
left precuneus		5.66	-12	-50	38
left inferior temporal gyrus	375	6.28	-52	-4	-32
left temporal pole		5.47	-48	10	-30
left middle temporal gyrus		5	-56	8	-26
right inferior parietal lobule	1546	6.2	60	-40	46
right supramarginal gyrus		5.29	64	-36	36
right angular gyrus		5.06	50	-66	28
left middle temporal gyrus	1204	5.53	-58	-56	22
left inferior parietal lobule		5.37	-60	-46	40
left supramarginal gyrus		5.14	-62	-46	32
right temporal pole	231	5.05	40	16	-34
right middle temporal gyrus		4.41	52	4	-30
right temporal pole		4.4	44	6	-38
right medial frontal gyrus	1135	4.86	6	56	4
left medial frontal gyrus		4.79	-4	54	8
right medial orbitofrontal cortex		4.62	8	50	-4
<b>Individual &gt; Arbitrary</b>					
left angular gyrus	736	5.6	-54	-58	26
left inferior parietal lobule		5.51	-52	-60	40
left inferior parietal lobule		4.57	-48	-74	36
right inferior temporal gyrus	122	5.13	56	0	-34
right inferior temporal gyrus		4.14	48	6	-34
left middle temporal gyrus	196	4.34	-54	-20	-20
left inferior temporal gyrus		4.22	-50	-4	-34
left middle temporal gyrus		4.08	-60	2	-24

**Table 4.1:** Whole-brain results for the comparisons between the collaborative, individual and arbitrary conditions. (*continued*)

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>Collaborative &gt; Individual</b>					
left putamen	279	6.07	-22	10	0
right midcingulate cortex	5056	5.73	14	-30	40
left superior parietal lobule		5.19	-16	-60	58
right precuneus		4.88	6	-52	56
right supramarginal gyrus	514	5.61	60	-42	36
right supramarginal gyrus		4.8	66	-42	30
right superior temporal gyrus		3.83	66	-38	20
left middle frontal gyrus	121	5.58	-28	36	30
right middle frontal gyrus	162	5.11	32	38	28
anterior cingulate cortex		3.49	20	30	28
anterior cingulate cortex	42	4.71	14	20	26
right medial frontal gyrus / supplementary motor area	173	4.56	2	0	48
left superior frontal gyrus	72	4.24	-16	-10	70
left inferior frontal gyrus	42	4.24	-38	8	10
right medial frontal gyrus	49	3.89	8	56	4
right medial frontal gyrus / supplementary motor area	63	3.88	10	-14	72
right precentral gyrus		3.58	16	-24	76
right lingual gyrus	42	3.74	18	-56	-10
<b>Arbitrary &gt; Individual</b>					
left inferior occipital gyrus	402	6.78	-54	-66	-16
left cerebellum		4.13	-46	-74	-22
		3.92	-30	-90	-20
left caudate nucleus	1199	6.39	-12	14	-6
left thalamus		5.26	-2	-16	12
right caudate nucleus		4.83	14	16	-4
left anterior cingulate cortex	1508	5.84	-6	10	28
left anterior cingulate cortex		5.38	-2	34	22
right anterior cingulate cortex		5.18	4	28	28
left cerebellum	289	5.57	-6	-56	-20
left cerebellum		4.91	-8	-72	-24
white matter	691	5.2	-12	-14	32
left insula		5.18	-32	18	-4

**Table 4.1:** Whole-brain results for the comparisons between the collaborative, individual and arbitrary conditions. (*continued*)

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
left insula		4.89	-34	8	20
white matter	98	5.17	-10	-32	28
white matter	241	5.11	-18	-12	-10
left hippocampus		4.86	-30	-22	-8
left amygdala		3.82	-22	-2	-10
left medial frontal gyrus / supplementary motor area	346	4.92	-8	-4	74
right medial frontal gyrus / supplementary motor area		4.89	10	-4	72
left medial frontal gyrus / supplementary motor area		3.74	0	-6	68
right middle frontal gyrus	108	4.9	32	38	28
left cerebellum	80	4.83	-30	-60	-34
right insula	950	4.8	32	24	4
right insula		4.54	52	14	-6
right insula		4.47	30	24	-4
left superior parietal lobule	1315	4.54	-30	-64	44
right calcarine gyrus		4.38	16	-98	4
left calcarine gyrus		4.37	2	-88	-10
white matter	52	4.5	-28	-36	28
right caudate nucleus	82	4.48	20	-20	24
right thalamus		4.18	18	-18	14
left precentral gyrus	121	4.46	-36	0	36
		3.86	-26	2	40
left precentral gyrus		3.62	-46	2	34
left cerebellum	55	4.36	-2	-46	-22
left postcentral gyrus	116	4.33	-48	-14	44
left postcentral gyrus		4.13	-42	-18	36
right rolandic operculum	42	4.22	66	0	12
left cerebellum	73	4.16	-48	-58	-34
brainstem	83	4.12	0	-14	-12
		3.75	2	-24	-12
		3.57	-8	-14	-4
right precuneus	81	4	18	-66	24
right superior occipital gyrus		3.97	26	-64	28
right calcarine gyrus		3.53	12	-78	12
	48	3.87	0	-36	4

## Discussion

In this study, participants generated labels for abstract figures together with another person in a collaborative communication task (collaborative condition) and by themselves (individual condition), and were given pre-determined, unrelated labels to learn by themselves (arbitrary condition). They then retrieved these labels during a communication task in the MRI scanner. The analysis of the fMRI data revealed two main findings. First and most importantly, we show that collaboratively learned labels are neurally distinguishable from individually learned labels mainly in brain areas related to social cognition. Second, we show that the retrieval of self-generated labels as compared to unrelated, arbitrary labels engages semantic processing areas, while retrieving arbitrary word-picture associations is more hippocampus-dependent.

### **Collaboratively encoded labels are neurally distinguishable from individually encoded labels**

The most direct test of the effect of collaborative encoding on memory retrieval is the comparison between the retrieval of collaboratively and individually learned labels. Labels learned in both of these conditions were self-generated, so the crucial difference between them is whether the labels were generated in a collaborative or an individual context.

The results of this comparison include the right angular gyrus<sup>1</sup>, the medial prefrontal cortex and the precuneus. The temporoparietal junction, the medial prefrontal cortex and the precuneus are considered to be key areas in the mentalizing or theory of mind network (Frith & Frith, 2006; Schurz et al., 2014; Van Overwalle, 2009), which is activated when people make inferences about other people's mental states. The mentalizing network has previously been found to be recruited when people perform communicative actions (e.g., Sassa et al., 2007; Willems et al., 2010) and during the encoding and retrieval of social information (e.g., Mitchell, Heatherton, & Macrae, 2002; Mitchell, 2004). For example, Mitchell (2004) found that subsequent memory correlates with activity during encoding in the medial prefrontal cortex during a social orienting task. Yet, our study is the first to show that collaborating with another person during the encoding of memories results in increased activity in the mentalizing network during their retrieval. Like the results of Mitchell and colleagues, this finding supports the idea of encoding specificity, which predicts that memories are encoded with specific cues related to the context in which they are encoded (Tulving & Thomson, 1973).

One important question is why we find increased activity in brain regions involved in mentalizing during the retrieval of collaboratively generated labels. Are these effects

---

<sup>1</sup> The angular gyrus is often referred to as the temporoparietal junction in the social cognition literature.

partner-specific, or more generally associated with collaborative encoding? Future research will have to show whether the mentalizing network is activated to the same extent when collaboratively generated labels are retrieved for a different addressee or in a non-communicative context. While we also observed behavioral differences between the labels learned in the collaborative and individual contexts at retrieval, it seems unlikely that these behavioral differences can explain the observed neural differences. For example, although the collaboratively generated labels contained more words, we did not observe any neural differences in the classical left-hemispheric language network.

### **The retrieval of self-generated labels engages semantic processing areas**

To test the effect of studying self-generated labels, we directly compared the retrieval of labels learned in the individual and arbitrary tasks. This comparison revealed clusters in the left angular gyrus and the bilateral temporal poles. We also found considerable overlap between the individual > arbitrary and collaborative > arbitrary contrasts in these areas (Figure 4.5, bottom row). In the collaborative and individual tasks, participants were free to come up with labels for the abstract figures. They probably tried to associate the abstract figures with their existing conceptual knowledge of the world. The temporal and angular clusters we find here are therefore likely the result of retrieving semantic associations between the pictures and labels (Binder et al., 2009; Patterson, Nestor, & Rogers, 2007; Price, 2012).

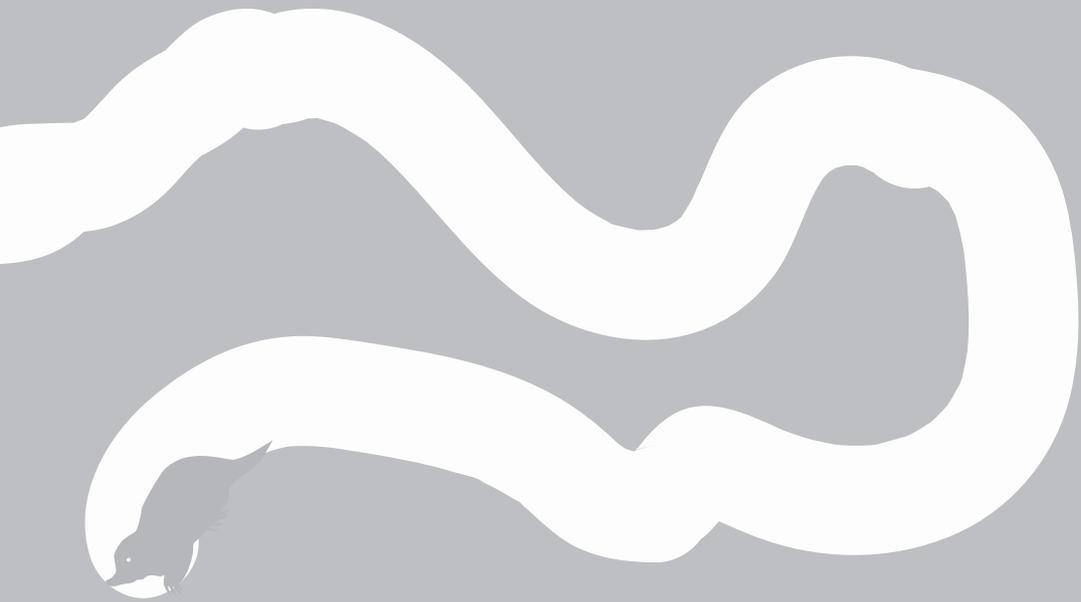
We only found a cluster in the hippocampus in the comparison between the arbitrary and the individual conditions. However, the hippocampus is probably to different degrees involved in the retrieval of labels generated in all conditions, with the arbitrary condition being most hippocampus-dependent and the individual condition being least hippocampus-dependent. The involvement of the hippocampus in collaborative learning also became evident in follow-up research to the original study by Duff and colleagues (2006). While patients with hippocampal amnesia are able to learn linguistic labels in a communicative task, they are impaired in their use of definite reference in this task (Duff, Gupta, Hengst, Tranel, & Cohen, 2011) and have difficulties establishing linguistic labels for highly similar pictures (Duff et al., 2012). It remains to be tested whether patients with hippocampal amnesia would perform similarly to healthy controls in the individual task we used, and whether they would show the same partner-specific effects as observed in healthy controls (Metzing & Brennan, 2003).

## **Collaboratively generated labels are better remembered**

The behavioral results show that collaboratively generated labels were remembered better than labels learned in the individual and arbitrary conditions. This is an interesting finding given that the collaboratively generated labels were also on average the longest. Based on the literature on collaborative encoding and retrieval (Barber, Rajaram, & Aron, 2010; Basden, Basden, Bryner, & Thomas, 1997; Weldon & Bellinger, 1997), one may have expected poorer recall performance in the collaborative compared to the individual condition. One likely explanation for the collaborative benefit we find here is that the collaborative task induced more elaborative processing. Over the course of the interactions during the collaborative task, directors may have formed richer memory representations than in the individual task. For example, if the matcher fails to select the right picture based on the director's initial description, the director has to come up with a novel description of the picture or provide additional detail, thus creating additional cue-target associations and facilitating retrieval (similar to the elaboration account of the testing effect proposed by Carpenter, 2009). Such prompts for elaboration are not present in the individual task. In addition, unlike previous studies on collaborative encoding, our collaborative task required participants to jointly come up with and agree on a set of labels to allow for efficient communication. This may have led participants to come up with coherent labels, which promote efficient collaborative encoding (Barber, Rajaram & Paneerselvam, 2012).

In conclusion, the present study compared the retrieval of labels learned in collaborative, individual and arbitrary contexts. Our results show that the retrieval of collaboratively generated labels as compared to individually learned labels engages brain regions involved in mentalizing. This study is the first to show that collaboration during encoding can affect the neural networks involved in retrieval.





The role of the mentalizing  
network in the establishment  
of conceptual pacts



## **Abstract**

Language is a joint action that requires coordination between interlocutors. One way in which this happens is through the establishment of conceptual pacts, i.e. temporary and flexible agreements about how to refer to something. In a previous fMRI study, we found that the retrieval of such jointly established linguistic labels recruits brain regions involved in mentalizing, including the medial prefrontal cortex, the right temporoparietal junction and the precuneus. In this study, we further investigated the role of mentalizing in the establishment of conceptual pacts by studying the neural representation of jointly established linguistic labels over the course of repeated references. Participants generated descriptions for abstract figures either with another participant (communicative blocks) or by themselves (individual blocks). We found that brain regions involved in mentalizing were most strongly recruited during the initial stages of jointly establishing linguistic labels and following misunderstandings between the participants. In addition, we found evidence that the establishment of conceptual pacts at least partly relies on domain-general episodic memory processes.

Based on:

Vanlangendonck, F., Willems, R. M., & Hagoort, P. (in preparation). The role of the mentalizing network in the establishment of conceptual pacts.

## Introduction

Language use is a form of joint action (Clark, 1996). A key feature of joint actions is that they require participants to coordinate their actions (Clark, 1996; Sebanz, Bekkering, & Knoblich, 2006). When people talk to each other, they coordinate turn-taking (Sacks, Schegloff, & Jefferson, 1974), spatial perspectives (Schober, 1993), visual attention (Richardson, Dale, & Kirkham, 2007) and meaning (Brennan & Clark, 1996). In this study, we investigated how the process of coordinating meaning shapes the neural representations of linguistic labels that are jointly established by communicators.

What does coordination of meaning look like? Imagine the following conversation between two friends looking at the store window of a shoe store:

- *What do you think of those pumps?*
- *The pointy ones?*
- *Yes, next to the purple shoes.*
- *They might look good with the dress we saw earlier.*

Over the course of this short interaction, the friends have jointly established a label for the shoes (“the pointy ones”) and have thus established a *conceptual pact*, i.e. a temporary, flexible agreement about how to conceptualize something (Brennan & Clark, 1996). Once a conceptual pact has been established, speakers are likely to use it again in later references (Brennan & Clark, 1996). The establishment of conceptual pacts has been studied extensively using referential communication tasks such as the tangram task (Clark & Wilkes-Gibbs, 1986), in which a director instructs a matcher to reorder the same set of abstract tangram figures multiple times. The labels speakers use tend to become shorter and more efficient with each repetition of the task (Clark & Wilkes-Gibbs, 1986).

Despite the extensive behavioral literature, the cognitive and neural mechanisms that underlie the establishment of conceptual pacts are still poorly understood. There is evidence that conceptual pacts are partner-specific (e.g., Brennan & Clark, 1996; Horton & Gerrig, 2005; Metzger & Brennan, 2003), but it is unclear how such partner-specific information is stored in memory. Horton & Gerrig (2005) proposed that domain-general episodic memory processes could support the associations between conceptual pacts and conversation partners. In this view, conversation partners create rich episodic memory traces of their interactions, thus linking interlocutors and conceptual pacts in memory. The presence of an interlocutor can then lead to the automatic retrieval of related information.

However, there is also evidence that learning labels in a communicative as compared to an individual setting may at least to some extent rely on neurally distinguishable resources.

One interesting line of research (Duff, Hengst, Tranel, & Cohen, 2006) showed that patients with hippocampal amnesia can successfully establish conceptual pacts over a series of interactions, while they are unable to learn arbitrary relations between labels and pictures in an individual paired-associate learning task. In a previous fMRI study, we directly compared the retrieval of conceptual pacts as opposed to individually encoded linguistic labels. We found that the retrieval of conceptual pacts recruits brain regions involved in mentalizing, i.e. people's ability to mentally put themselves in someone else's shoes (Amodio & Frith, 2006; Frith & Frith, 2006; Frith & Frith, 2010; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Van Overwalle, 2009). These findings suggest that conceptual pacts are neurally distinguishable from individually learned labels in brain areas involved in mentalizing and that the process of jointly establishing a linguistic label can affect the brain networks involved at retrieval.

In the current study, we aimed to investigate how brain regions involved in mentalizing contribute to the *establishment* of conceptual pacts. Participants in the experiment performed a communicative task based on the tangram task and an individual control task in the MRI scanner. In the communicative task, participants generated linguistic labels for a set of tangram pictures, while another participant outside the scanner tried to click on the described pictures. Participants in the scanner received direct feedback on the other participant's performance, allowing them to adjust their descriptions until they established a set of shared linguistic labels. In the individual task, participants in the MRI scanner came up with linguistic labels for a different set of tangram figures by themselves. Participants in the MRI scanner also completed a theory of mind localizer to independently localize brain regions involved in mentalizing. This design allowed us to investigate the recruitment of brain regions involved in mentalizing over the course of multiple rounds of the tangram task. By directly comparing the communicative and individual tasks, we could control for the effects of repeated viewing and naming of the figures.

Based on our previous study, we expected that communicative as compared to individual encoding of linguistic labels would engage parts of the mentalizing network. We were especially interested in studying how representations of conceptual pacts change as they are established and whether this would lead to increased or decreased engagement of the mentalizing network. In line with the principle of least collaborative effort (Clark & Wilkes-Gibbs, 1986), one may expect that interlocutors mostly use mentalizing during the initial stages of establishing a conceptual pact. On the other hand, our previous study suggests that the mentalizing network is still consistently involved at retrieval even after several rounds of the tangram task. At the whole-brain level, we were especially interested in the recruitment of memory structures involved in episodic memory encoding and retrieval to see whether we could find support for the involvement of episodic memory processes as proposed by Horton & Gerrig (2005). In addition to studying the establishment of conceptual pacts, we

were interested in the influence of feedback from the matcher on the coordination process. We therefore compared first attempts to describe the figures (without feedback from the matcher) to later attempts after having received feedback from the matcher. We expected that the effect of communicative context may be maximal in the case of a misunderstanding, when mentalizing may be required to adjust your response based on your addressee's feedback.

## Method

### Participants

72 native Dutch speakers (36 pairs) participated in the experiment. They all had normal or corrected-to-normal vision and no history of neurological disease. Participants gave written informed consent before the start of the experiment and received course credit or monetary compensation for their participation. Data from six pairs were excluded (two pairs due to excessive motion by the participant in the MRI scanner, one pair due to technical problems with the sound recordings and three pairs due to artifacts during data acquisition). The results of the remaining thirty pairs (MRI: 8 men; 22.73 years old, range 18-36 years; behavioral: 5 men, 22.13 years old, range 18-31 years) are reported below. Four MRI participants did not complete the theory of mind localizer.

### Procedure

The experiment consisted of a picture description task and a theory of mind localizer. Before the start of the experiment, participants were randomly assigned to the roles of director or matcher for the duration of the experiment. The director performed the picture description task in the MRI scanner using a noise-cancelling microphone and a button box. The matcher was seated in the MRI lab wearing headphones and responded using a mouse. The directors' verbal responses were played to the matcher in real time and were recorded for offline coding.

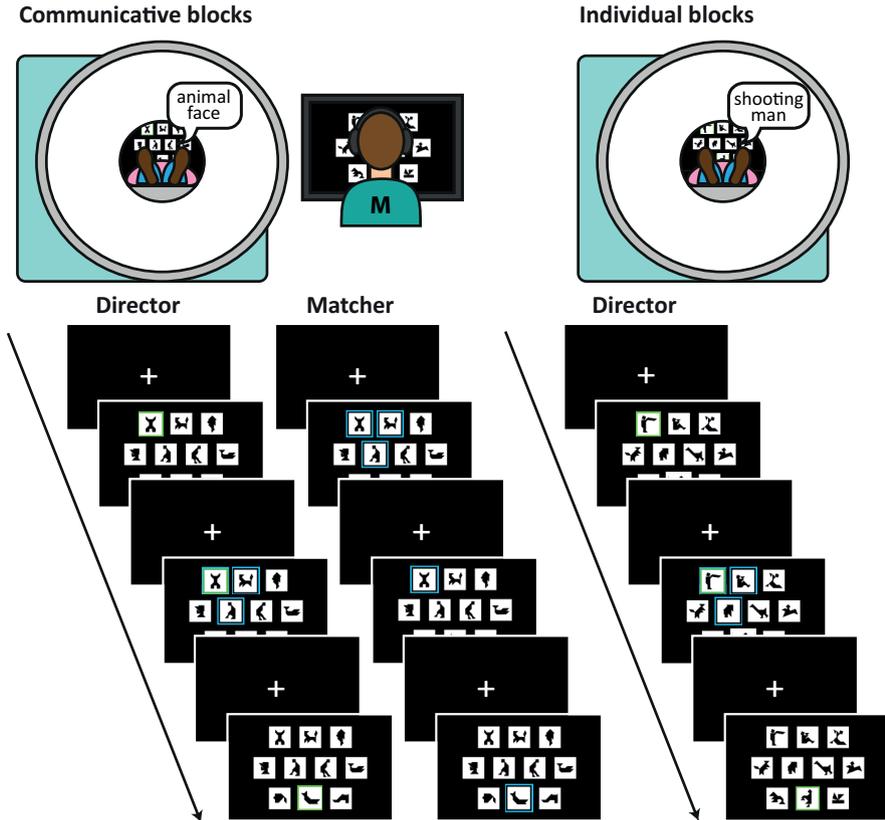
The picture description task consisted of three communicative blocks and three individual blocks. Each block featured a new set of ten tangram pictures. During the communicative blocks, the directors and matchers were presented with the same set of ten tangram figures. On each trial, directors were cued to describe the picture around which a green square appeared using a maximum of two words. They came up with a description for the picture, pressed a button when they were ready to start speaking, and then said their description out loud. The matchers could then select one or multiple pictures depending on how certain they were of what they thought the director had described. The matchers saw blue squares appear around the pictures they clicked on. If the matchers clicked on the described picture, and this was the only picture they clicked on, the trial ended. The directors were then cued

to describe the next tangram picture. In all other cases (multiple pictures, no pictures or wrong picture selected), the directors were cued (with a green square) to describe the same picture again and received feedback about which pictures the matcher had selected based on their previous description (blue squares). During these repeated attempts, the message “same picture” appeared on the matchers’ screen so that they knew that the director was describing the same picture again. After four failed attempts to describe a specific picture, the task moved on to the next target picture.

The individual blocks were visually very similar to the communicative blocks, but the matchers did not participate in these blocks. Like in the communicative blocks, the directors saw ten tangram figures on screen at a time and were cued to describe the picture around which a green square appeared using at most two words. The directors knew that the matchers did not participate in the individual blocks and were instructed to come up with labels for the pictures based on what they saw in the pictures. They came up with a description for the target picture, pressed a button when they were ready to start speaking, and then said their description out loud. On some trials, the directors were cued to describe the same picture multiple times in a row. In that case, they were instructed to change their previous description (e.g., by modifying their previous description or generating a new label). Occasionally, blue squares could appear around some of the pictures. Directors were told that the repetition of target pictures and the appearance of blue squares around certain pictures were determined by the computer to make the individual and communicative blocks as similar as possible, and that they did not depend on the quality of the directors’ responses.

Before the start of the experiment, participants practiced the picture description task outside the scanner. They completed one communicative and one individual block of the task using sets of tangram pictures that were not used in the MRI experiment. During the MRI phase, the communicative and individual blocks alternated, and the experiment always started with a communicative block. This allowed us to carefully match the number of repetitions and the visual presentation of the stimuli between communicative and individual blocks. The number of repetitions per picture and the kind of feedback the director received in the communicative blocks were recorded and used to determine the repetitions and appearance of blue squares in the individual blocks. For example, if the third trial in the first communicative block was repeated and the matcher had clicked on three pictures, then the third trial in the next individual block was repeated and the director saw blue squares in the same locations as the pictures the matcher had clicked in the communicative block (see Figure 5.1).

Each block of the task featured a new set of ten pictures and consisted of three rounds. In each round, the director described each picture at least once (depending on the number of repetitions). The locations of the pictures and the order in which they were cued were randomized before each communicative block. The following individual block used the same picture randomization. Trials for the director lasted 7 s and were preceded by a 3-5 s jittered inter-trial interval. Trials for the matcher lasted 9.5 s (7 s of the director's trial + 2.5 s of the director's ITI) to give them enough time to click on pictures.



**Figure 5.1:** The experimental set-up and example trials from the picture description task. In the communicative blocks, the director describes the target pictures for the matcher outside the scanner. On each trial in the communicative blocks, the director describes a target picture (in green) and the matcher tries to click on the described picture based on the director's response. In the communicative example trial sequence (left), the matcher initially clicks on three pictures. The director then receives the feedback from the matcher (in blue) and tries to describe the next target picture again. When the matcher selects the correct picture, the director is cued to describe the next target picture and the matcher tries to click on this picture. In the individual blocks, the director describes the target pictures for him- or herself. The trial sequence in the individual blocks (right) is matched to the trial sequence in the communicative blocks, but uses a different set of tangram pictures.

After the picture description task, participants in the MRI scanner completed a theory of mind localizer task (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2010; Saxe & Kanwisher, 2003; Van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). During this task, they were presented with twenty stories that required them to represent false content. In half of the stories the false content concerned the physical state of an object (false photograph); in the other half of the stories it concerned another person's belief (false belief). Each story was presented for ten seconds, after which participants were given a statement about the story to judge. They had to respond to the statements with a button press within five seconds. A variable inter-trial interval of 4000-8000 ms preceded each trial.

## Materials

The stimuli consisted of sixty abstract tangram figures, which were adapted from Read (1965). The stimulus set was initially created for and previously used in another experiment (chapter 4). We divided the sixty pictures into six sets of ten pictures based on difficulty ratings and the predominant semantic category (animal, person or object) of the responses obtained in a pretest. The order of the picture sets and the pairing between picture sets and blocks were counterbalanced.

## Data acquisition and analysis

Participants were scanned in a Siemens 3T Prisma MRI scanner using a 32-channel head coil. Functional images were acquired using a T2\*-weighted gradient multi-echo planar imaging sequence (TR = 2240 ms; TE1 at 9 ms, TE2 at 19.1 ms, TE3 at 29.14 ms, TE4 at 39.21 ms, TE5 at 49.28 ms; 32 slices; ascending slice order; 3 mm slice thickness; 0.56 mm slice gap; 64 x 64 matrix size; 224 x 224 mm FOV; 90° flip angle and 3.5 x 3.5 x 3 mm voxel size). In addition, T1-weighted anatomical scans with 1 mm isotropic resolution were acquired (TR = 2300 ms; TE = 3.03 ms; 8° flip angle; 256 x 256 x 192 mm FOV).

We acquired 35 pre-scans before the start of each task in the MRI scanner. These scans were used to calculate the optimal weighting of the five echoes, and this weighting matrix was applied to the remaining functional scans (Poser, Versluis, Hoogduin, & Norris, 2006). Preprocessing was done in SPM12 (Statistical Parametric Mapping, [www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). The preprocessing of the functional images consisted of slice timing correction to the onset of the middle slice, coregistration of the functional images to the T1 based on the subject-mean functional image, normalization to MNI space (resulting voxel size 2 x 2 x 2 mm) and spatial smoothing using a 3-dimensional isotropic Gaussian smoothing kernel (full-width half-maximum = 8 mm).

The speakers' verbal responses were transcribed and coded by comparing the responses both within and across rounds. First, we wanted to test whether directors used consistent responses over multiple rounds of the picture description task, i.e. whether they repeated their previous responses when they were cued to describe the same picture in a later round. In order to test this, we compared the responses from the second and third rounds to all preceding responses. Only responses that were identical or that at least partly overlapped in form with previous responses were considered to be consistent (e.g., *bird* vs. *large bird*). Second, we coded whether responses were a first attempt to describe a picture (i.e., initial response in a round) or a repetition (i.e., attempts after the initial response per round).

**Consistency model.** We created two types of first-level general linear models to analyze the results of the picture description task. The first model allowed us to look at how linguistic labels are established over the course of repeated references. We focused on consistent responses (see coding above), as speakers were expected to reuse previously established conceptual pacts in later references. Consistent hence means consistent with the descriptions in previous rounds. This model allowed us to directly compare consistent responses in the different rounds of the communicative and individual blocks. We included eight variables in these models: first round responses in the communicative blocks, first round responses in the individual blocks<sup>1</sup>, second round consistent responses in the communicative blocks, second round consistent responses in the individual blocks, third round consistent responses in the communicative blocks, third round consistent responses in the individual blocks, speaking, and responses of no interest. The first six regressors were stick functions time-locked to picture onset. We included the planning duration associated with each response as a linear parametric modulator. The speaking regressor was modeled as a stick function and we included the number of words per response as a linear parametric modulator. The condition of no interest regressor was modeled as a boxcar function (7 s) and included inconsistent responses from the second and third rounds (i.e., responses without at least partial overlap with previous responses), as well responses without a button press or a verbal response and responses that contained speech unrelated to the task.

**Repetition model.** The second type of model allowed us to test the effect of communicative context on first attempts to describe a picture as well as on repetitions after directors received feedback from the matcher. This model allowed us to directly compare first attempts and repetitions between the communicative and individual blocks, collapsed across rounds. We included six variables in these models: first attempts in the communicative blocks, first attempts in the individual blocks, repetitions in the communicative blocks, repetitions in

<sup>1</sup> We could not compare the responses from the first round to previous responses, so we included all first round responses in the first round regressors (except responses without a verbal response or a button press and responses that contained speech unrelated to the task).

the individual blocks, speaking, and responses of no interest. The first four regressors were stick functions time-locked to picture onset. We included the planning duration associated with each response as a linear parametric modulator. The speaking regressor was modeled as a stick function and we included the number of words per response as a linear parametric modulator. The condition of no interest regressor was modeled as a boxcar function (7 s) and included responses without a button press or speech, as well as responses that contained speech unrelated to the task.

In addition to the models used to analyze the picture description data, we created first-level general linear models to analyze the theory of mind localizer. These statistical models included four regressors: false belief stories, false photograph stories, false belief statements and false photograph statements. We used boxcar functions to model the durations (10 s for stories; 5 s for statements).

For all models, regressors were convolved with the hemodynamic response function. In addition, we included 24 nuisance regressors: the six realignment parameters, their square, their first derivative, and the realignment parameters used to realign the previous volume. Individual t-contrasts were created and used in second-level random-effect analyses. Group analyses were performed using one-sample t-tests. Whole-brain results were corrected for multiple comparisons by combining a  $p < 0.001$  voxel-level threshold with a cluster extent threshold determined by means of a Monte Carlo simulation with 2500 iterations, after estimation of the smoothness of the data (Slotnick, Moo, Segal, & Hart, 2003). This revealed that clusters of 46 voxels or larger indicated statistically significant effects at the  $p < 0.05$  level, corrected for multiple comparisons.

We ran region of interest (ROI) analyses using the brain regions extracted from the contrast between false belief and false photograph statements in the theory of mind localizer. Based on the results of our previous fMRI experiment, we were especially interested in the following ROIs: the medial prefrontal cortex (mPFC), the left temporoparietal junction (TPJ), the right TPJ and the precuneus. Unlike the left TPJ, the right TPJ cluster extracted from the theory of mind localizer was part of a large cluster that extended to the right temporal pole. We therefore only included the part of this cluster that was posterior to the most anterior coordinate of the left TPJ cluster ( $y = -44$ ) in the right TPJ ROI. Parameter estimates based on these ROIs were computed using MarsBar (Brett, Anton, Valabregue, & Poline, 2002) for a number of contrasts based on the consistency model (communicative first round > individual first round, communicative second round > individual second round, communicative third round > individual third round) and the repetition model (communicative first attempt > individual first attempt, communicative repetition > individual repetition).

## Results

### Behavioral results

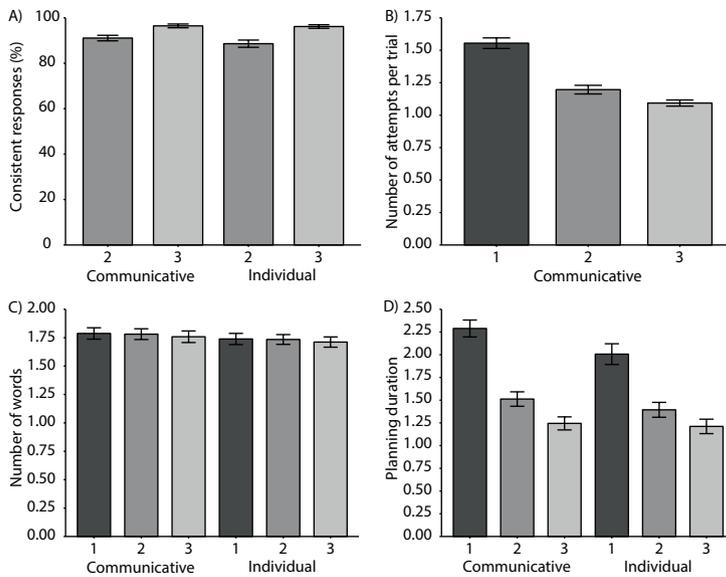
We compared the communicative and individual blocks on a number of behavioral measures. Based on the coding of the responses we looked at the degree of consistency in responses in the second and third rounds and the number of attempts per trial per round of the picture description task. In addition, we counted the mean number of words per response and calculated the planning durations per trial, i.e. the duration between stimulus onset and the director's button press. Greenhouse-Geisser correction was applied when the sphericity assumption was violated and Holm-Bonferroni correction for multiple comparisons was used for all reported post-hoc tests. The results are summarized in Figure 5.2.

**Consistency.** We ran a 2 x 2 repeated-measures ANOVA comparing the percentage of consistent responses with block (communicative or individual) and round (second or third round) as within-subject factors. Participants produced more consistent responses with every round of the task,  $F(1,29) = 50.07, p < 0.001$ . We did not find a significant main effect of block ( $F(1,29) = 2.64, p = 0.12$ ), nor a significant block x round interaction ( $F(1,29) = 1.56, p = 0.22$ ).

**Number of attempts.** We ran a repeated-measures ANOVA with round as within-subject factor. We did not include block as a factor, as the number of repetitions in the individual blocks was matched to the number of repetitions in the communicative blocks. We found a significant main effect of round on the number of attempts per trial,  $F(2,58) = 158.39, p < 0.001$ . Post-hoc paired t-tests revealed significantly more attempts in the first compared to the second round ( $t(29) = 11.97, p < 0.001$ ), and in the second round compared to the third round ( $t(29) = 5.97, p < 0.001$ ).

**Number of words.** We ran a 2 x 3 repeated-measures ANOVA comparing the percentage of consistent responses with block (communicative or individual) and round (first, second or third round) as within-subject factors. We did not find significant main effects (block:  $F(1,29) = 2.08, p = 0.16$ ; round:  $F(2,58) = 1.15, p = 0.33$ ), nor a significant block x round interaction ( $F(2,58) < 1, p = 1$ ).

**Planning durations.** We ran a 2 x 3 repeated-measures ANOVA comparing the percentage of consistent responses with block (communicative or individual) and round (first, second or third round) as within-subject factors. We found a significant main effect of block ( $F(1,29) = 14.32, p < 0.001$ ), a main effect of round ( $F(2,58) = 203.86, p < 0.001$ ) and a significant block x round interaction effect ( $F(2,58) = 9.02, p < 0.01$ ). Post-hoc paired t-tests revealed significantly longer planning durations in the first collaborative round relative to the first individual round ( $t(29) = 4.60, p < 0.001$ ), but no significant differences between the second rounds ( $t(29) = 2.10, p = 0.09$ ) and the third rounds of the block ( $t(29) = 1.06, p = 0.30$ ).



**Figure 5.2:** Behavioral results: A) the percentage of responses that are consistent with previously given responses in the second and third rounds of the communicative and individual blocks (first round responses are not included, as there were no responses from previous rounds to compare them to), B) the mean number of attempts per trial in the communicative blocks (which were matched in the individual blocks), C) the mean number of words per response given, and D) the mean planning duration per response, as measured from trial onset until the director's button press response. Error bars indicate standard error of the mean.

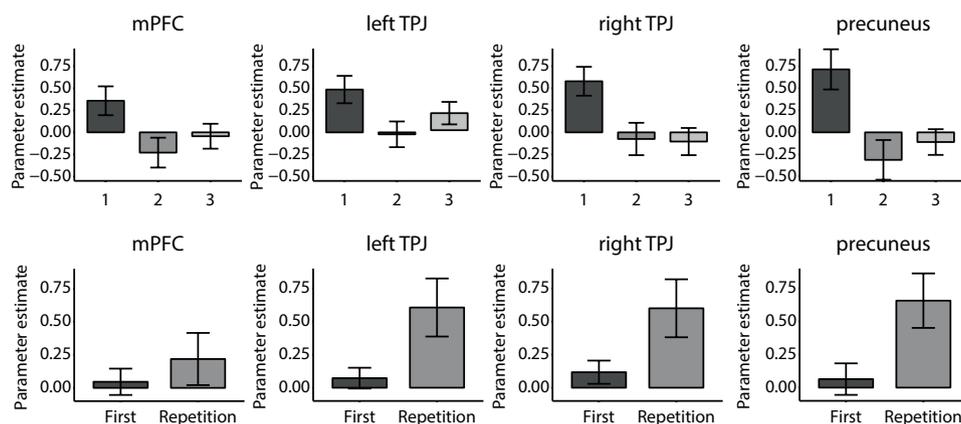
## ROI results

For the ROI analysis, we focused on a subset of the clusters found in the contrast between false belief and false photograph statements from the theory of mind localizer: the mPFC, the left TPJ, the right TPJ and the precuneus. For the consistency model, we looked at the direct comparisons between consistent responses in the communicative and individual blocks per round to investigate how the recruitment of mentalizing areas changes as speakers establish and reuse conceptual pacts. For the repetition model, we looked at the direct contrasts of the first attempts and the repetitions between the communicative and the individual blocks. This allowed us to test whether mentalizing areas are especially recruited following a misunderstanding in a communicative context. The parameter estimates obtained from the ROI analyses are shown in Figure 5.3.

**Consistency model.** We ran a repeated-measures ANOVA with round as within-subject factor on the parameter estimates obtained in the regions of interest using the communicative first round > individual first round, communicative second round > individual second round and communicative third round > individual third round contrasts. We found a significant main effect of round in all ROIs (mPFC:  $F(2,58) = 3.68$ ,  $p < 0.05$ ; left TPJ:  $F(2,58) = 5.74$ ,  $p < 0.01$ ;

right TPJ:  $F(2,58) = 5.06$ ,  $p < 0.01$ ; precuneus:  $F(2,58) = 6.55$ ,  $p < 0.01$ ). Planned comparisons revealed significant differences between the first and second round (mPFC:  $t(58) = 2.58$ ,  $p < 0.05$ ; left TPJ:  $t(58) = 3.38$ ,  $p < 0.01$ ; right TPJ:  $t(58) = 3.26$ ,  $p < 0.01$ ; precuneus:  $t(58) = 3.72$ ,  $p < 0.001$ ). A significant difference between the second and third round was only found in the left TPJ (left TPJ:  $t(58) = 2.42$ ,  $p < 0.05$ ; mPFC:  $t(58) = 0.56$ ,  $p = 0.58$ ; right TPJ:  $t(58) = 1.74$ ,  $p = 0.09$ ; precuneus:  $t(58) = 1.25$ ,  $p = 0.22$ ).

**Repetition model.** We ran paired t-tests to compare the parameter estimates obtained in the first attempts and the repetition trials in the regions of interest using the communicative first attempt > individual first attempt and communicative repetition > individual repetition contrasts. Parameter estimates were significantly higher for the repetition trials in all regions except the mPFC (mPFC:  $t(29) = 0.81$ ,  $p = 0.43$ ; left TPJ:  $t(29) = 2.35$ ,  $p < 0.05$ , right TPJ:  $t(29) = 2.13$ ,  $p < 0.05$ ; precuneus:  $t(29) = 2.57$ ,  $p < 0.05$ ).



**Figure 5.3:** Results of the region of interest analysis. The top row contains the results of the ROI analysis for the direct comparisons per round between the communicative and the individual blocks (consistency model). The bottom row contains the results for the direct comparisons between the communicative and individual first attempts and repetitions (repetition model). Error bars indicate standard error of the mean.

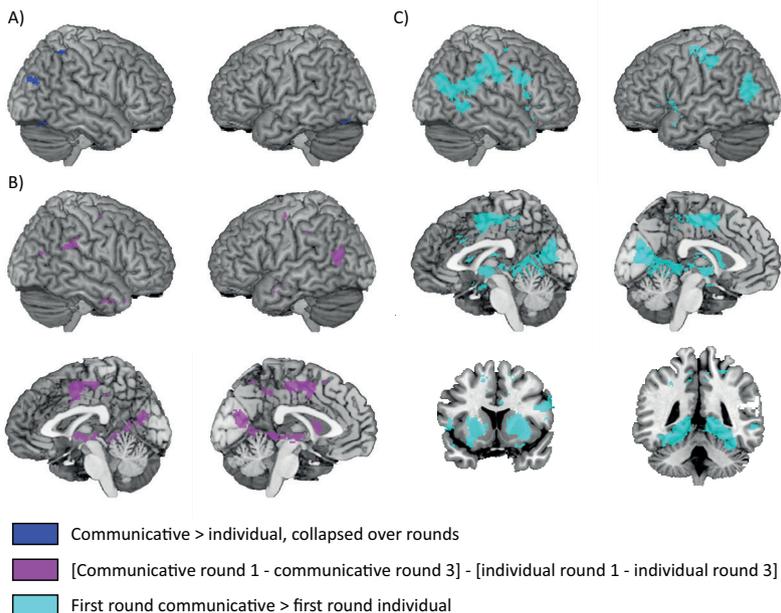
## Whole-brain results

In this study, we aimed to investigate how the neural representation of conceptual pacts changes over the course of repeated reference. By comparing the communicative blocks to the individual blocks, we could control for the effects of repeated viewing and naming of the pictures, thus allowing us to focus on the effect of coordinating meaning with a communication partner over rounds. For the whole-brain analysis, we first looked at the way the neural representation of conceptual pacts changes over rounds (*consistency model*). We therefore compared consistent responses between the communicative and individual

blocks. The results of these comparisons can be found in Table 5.1 and Figure 5.4. Given that the effect of communicative context may be maximal following a misunderstanding (i.e., when the matcher did not select the correct picture based on the director's description, and needs to describe the picture again), we then compared directors' first attempts and repeated attempts to describe pictures between the communicative and the individual blocks, collapsing over rounds (*repetition model*). The results of these contrasts are listed in Table 5.2 and Figure 5.5.

**Communicative > individual, collapsed over rounds.** The direct comparison of the communicative and the individual blocks, collapsed over rounds revealed a series of mostly right-lateralized clusters including the fusiform gyrus, the inferior parietal lobule and the inferior temporal gyrus.

**[Communicative round 1 - communicative round 3] - [individual round 1 - individual round 3].** We tested which brain regions showed a larger decrease in activity from the first to the third round in the communicative relative to the individual blocks. This directional interaction revealed a large cluster that included the bilateral parahippocampal gyri, hippocampi, fusiform gyri and lingual gyri. Another large and more anterior bilateral cluster included the thalamus, pallidum, putamen and amygdala. The remaining clusters are described in Table 5.1.



**Figure 5.4:** A) Brain areas showing greater activity in the communicative planning trials compared to the individual planning trials, collapsed over rounds (dark blue), B) brain regions showing activity for the directed block x round interaction effect (violet) and, C) brain regions showing greater activity for trials in the first communicative round relative to the first individual round (cyan). Below, we show the clusters in the putamen as well as in the hippocampus and parahippocampal gyrus.

**Communicative > individual per round.** The results of the direct comparison between the first rounds of the communicative and the individual blocks overlap to a considerable extent with the results of the previous contrast. The largest cluster extends from the bilateral insula, putamen, amygdala, pallidum, caudate and thalamus more posteriorly to the hippocampus, parahippocampal, fusiform and lingual gyri and the precuneus. In contrast, the comparison between the second rounds of the communicative and the individual blocks revealed no significant clusters, and the comparison of the third rounds only resulted in two cerebellar clusters.

**Table 5.1:** Whole-brain results of the contrasts based on the consistency model.

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>Communicative &gt; individual</b>					
right fusiform gyrus	291	5.27	34	-72	-18
right fusiform gyrus		3.74	36	-60	-20
right inferior parietal lobule	97	4.79	26	-52	54
right precuneus		3.71	14	-52	54
right inferior temporal gyrus	90	4.66	52	-44	-10
right inferior temporal gyrus		4.34	50	-44	-18
left fusiform gyrus	50	4.52	-36	-44	-24
right insula	46	4.48	34	20	-6
right middle occipital gyrus	67	4.33	46	-76	24
right middle occipital gyrus		4.27	40	-82	24
right middle occipital gyrus		3.51	32	-80	20
left fusiform gyrus	63	4.17	-34	-72	-16
right inferior parietal lobule	65	3.97	36	-42	50
right inferior parietal lobule		3.78	48	-52	54
<b>[Communicative round 1 - communicative round 3] - [individual round 1 - individual round 3]</b>					
left precentral gyrus	1323	6.88	-20	-14	58
left precentral gyrus		5.63	-32	-10	58
left supplementary motor area		5.03	-8	2	52
right parahippocampal gyrus	3488	6.61	22	-42	-8
left lingual gyrus		6.27	-16	-40	-2
left lingual gyrus		5.42	-22	-46	-4

**Table 5.1:** Whole-brain results of the contrasts based on the consistency model. (*continued*)

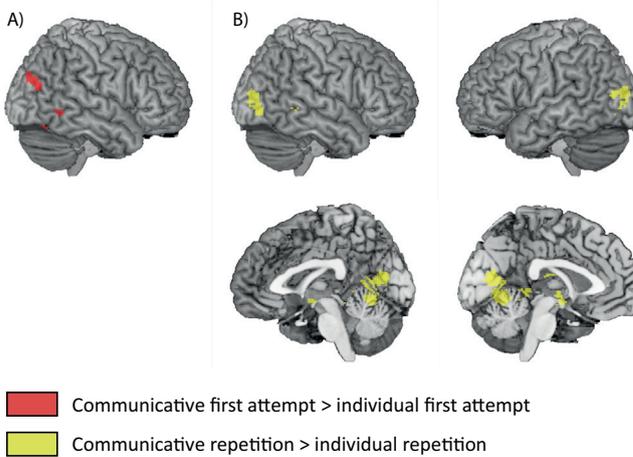
Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>[Communicative round 1 - communicative round 3] - [individual round 1 - individual round 3]</b>					
left middle temporal gyrus	420	5.85	-46	-62	14
left middle temporal gyrus		4.14	-44	-56	8
left middle temporal gyrus		3.71	-54	-68	22
left amygdala	2658	5.64	-34	0	-22
left thalamus		5.18	-14	-12	-2
right pallidum		5.15	16	6	-2
right inferior temporal gyrus	127	5.4	52	-16	-24
right temporal pole		4.46	48	10	-32
right temporal pole		4.24	36	16	-32
left inferior temporal gyrus	69	5.29	-40	-12	-30
right midcingulate cortex	88	4.82	12	-32	42
left precuneus	284	4.7	-10	-50	44
left precuneus		4.54	-12	-56	56
left midcingulate cortex		4.44	-6	-40	44
left inferior parietal lobule	255	4.63	-30	-40	42
left inferior parietal lobule		4.37	-44	-38	44
left inferior parietal lobule		3.63	-50	-30	44
right postcentral gyrus	199	4.49	36	-30	40
right postcentral gyrus		4.48	46	-26	40
right supramarginal gyrus		3.61	56	-28	38
right supramarginal gyrus	221	4.37	64	-36	26
right supramarginal gyrus		4.17	54	-22	20
right supramarginal gyrus		4.08	48	-30	24
right middle temporal gyrus	88	4.23	46	-58	10
right middle temporal gyrus		3.95	44	-66	22
<b>First round communicative &gt; first round individual</b>					
right precuneus	17815	7.51	18	-42	0
left precuneus		7.48	-14	-44	2
left insula		7.39	-34	0	-6
right precentral gyrus	1884	6.11	36	-6	58
right supplementary motor area		5.35	16	-8	56

**Table 5.1:** Whole-brain results of the contrasts based on the consistency model. (*continued*)

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>First round communicative &gt; first round individual</b>					
right supplementary motor area		5.25	8	-6	54
right middle temporal gyrus	198	5.76	64	-48	-2
right middle temporal gyrus		4.16	56	-46	-8
right middle temporal gyrus		3.86	50	-56	2
left precentral gyrus	1309	5.54	-42	-18	58
left superior frontal gyrus		5.53	-22	-8	58
left inferior parietal lobule		5.36	-50	-30	44
right precentral gyrus	460	5.47	60	10	28
right inferior frontal gyrus		4.34	54	16	10
right inferior frontal gyrus		4.31	52	6	22
left superior frontal gyrus	99	5.01	-10	32	48
left middle frontal gyrus		4.57	-22	30	38
left superior medial frontal gyrus		3.75	-6	34	40
left middle temporal gyrus	564	4.78	-46	-64	8
left middle occipital gyrus		4.54	-46	-80	16
left middle temporal gyrus		4.5	-42	-62	18
right midcingulate cortex	92	4.54	14	16	34
right anterior cingulate cortex		4.2	8	14	28
right anterior cingulate cortex		4.02	12	22	28
right thalamus	62	4.19	10	-26	2
		3.71	-2	-30	-2
left inferior frontal gyrus	77	4.11	-48	16	8
left superior temporal gyrus		4.05	-50	8	0
left superior temporal gyrus		3.99	-52	4	-10
<b>Second round communicative &gt; second round individual</b>					
no suprathreshold clusters					
<b>Third round communicative &gt; third round individual</b>					
right cerebellum	90	4.6	48	-70	-32
right cerebellum		3.98	38	-70	-22
left cerebellum	68	4.36	-42	-70	-26

**Communicative first attempt > individual first attempt.** We directly compared first attempts in the communicative and individual blocks, collapsed over rounds. This contrast revealed significant right-lateralized parietal, temporal, occipital and cerebellar clusters.

**Communicative repetition > individual first attempt.** The direct comparison between repetitions in the communicative and individual blocks revealed a relatively deep intraparietal cluster, as well as a large cluster including the bilateral lingual, calcarine, fusiform and parahippocampal gyri, the hippocampi and the cerebellum. The remaining clusters are described in Table 5.2.



**Figure 5.5:** A) Brain areas showing greater activity for the first attempts in the communicative blocks compared to the individual blocks, collapsed over rounds (red), B) brain regions showing greater activity for repeated attempts in the communicative blocks compared to the individual blocks, collapsed over rounds (yellow).

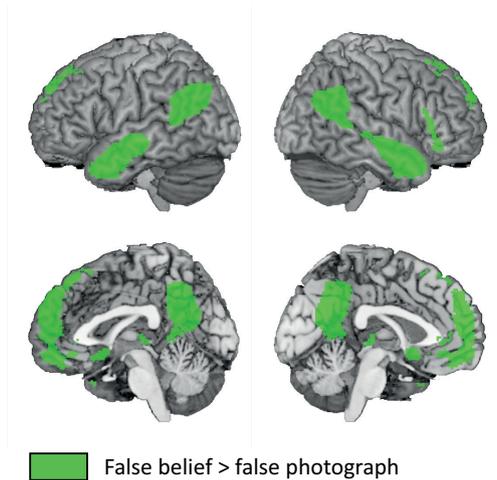
**Table 5.2:** Whole-brain results of the contrasts based on the repetition model.

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>Communicative first attempt &gt; individual first attempt</b>					
right inferior temporal gyrus	106	5.32	60	-54	-6
right inferior temporal gyrus		4.85	50	-44	-10
right inferior parietal lobule	49	4.85	26	-52	54
right postcentral gyrus		4.09	24	-42	56
right middle occipital gyrus	109	4.63	44	-78	26
right middle temporal gyrus		3.92	48	-66	14
right middle occipital gyrus		3.9	36	-82	32

**Table 5.2:** Whole-brain results of the contrasts based on the repetition model. (*continued*)

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>Communicative first attempt &gt; individual first attempt</b>					
right fusiform gyrus	141	4.54	36	-70	-20
right cerebellum		3.96	28	-72	-20
right cerebellum		3.72	24	-66	-24
<b>Communicative repetition &gt; individual repetition</b>					
left intraparietal sulcus	551	5.9	-30	-38	34
left intraparietal sulcus		5.3	-24	-48	34
left intraparietal sulcus		4.88	-34	-50	36
right fusiform gyrus	4235	5.88	28	-54	-10
right lingual gyrus		5.34	22	-58	-4
left lingual gyrus		4.94	-24	-50	-8
right intraparietal sulcus	123	4.89	38	-44	34
right intraparietal sulcus		3.7	44	-38	38
right middle temporal gyrus	79	4.78	46	-34	-4
right middle temporal gyrus		3.68	62	-42	-2
right superior temporal gyrus		3.62	46	-24	-4
right inferior occipital gyrus	118	4.64	46	-78	-4
right middle occipital gyrus		4.11	42	-84	10
right middle occipital gyrus		3.8	38	-88	2
left middle occipital gyrus	152	4.62	-36	-88	14
left middle occipital gyrus		4.28	-50	-76	10
left middle occipital gyrus		3.75	-36	-80	-2
left middle temporal gyrus	164	4.45	-50	-36	-2
left middle temporal gyrus		4.27	-48	-28	-2
	81	4.14	-2	-6	-10
left thalamus		3.8	-18	-8	2
left thalamus		3.7	-6	-12	-4
right amygdala	55	4.14	32	2	-12
right putamen		4.06	32	6	-4
left superior parietal lobule	68	3.95	-22	-58	56
left superior parietal lobule		3.85	-28	-48	50

**Theory of mind localizer: false belief > false photograph.** The contrast between false belief and false photograph statements in the theory of mind localizer revealed an activation pattern in line with previous studies (Dodell-Feder et al., 2010; Saxe & Kanwisher, 2003; van Ackeren et al., 2012), including bilateral clusters in the medial prefrontal cortex, temporal poles, middle temporal gyri, angular gyri and the precuneus (Figure 5.6, Table 5.3). We used the results of this contrast to determine the areas for the ROI analysis.



**Figure 5.6:** A) Brain areas showing greater activity for the false belief statements relative to the false photograph statements (green).

## Discussion

In this study, we investigated how the process of coordinating meaning shapes the neural representations of conceptual pacts. We were especially interested in how the recruitment of brain regions involved in mentalizing changes over the course of repeated references. Participants generated labels for abstract figures together with another person (communicative blocks) or by themselves (individual blocks). The ROI analysis revealed that the mentalizing network was most strongly engaged during the initial stages of establishing conceptual pacts and when multiple attempts were needed to describe a figure in the communicative blocks. The whole-brain analysis revealed neural differences in parietal, occipital and temporal brain regions between the communicative and the individual blocks. Similar to the ROI results, these neural differences were most pronounced during the initial stages of establishing conceptual pacts and following misunderstandings between the participants in the communicative blocks.

**Table 5.3:** Whole-brain results of comparison between the false belief and false photograph statements from the theory of mind localizer.

Brain region	Cluster extent (voxels)	T value	MNI coordinates		
			x	y	z
<b>False belief &gt; false photograph</b>					
right temporal pole	4508	12.51	54	10	-32
right middle temporal gyrus		10.44	56	2	-20
right middle temporal gyrus		10.29	54	-8	-18
left angular gyrus	2304	9.85	-44	-64	26
left angular gyrus		9.06	-46	-54	24
left middle occipital gyrus		7.95	-40	-72	32
left middle temporal gyrus	2025	9.7	-60	-16	-16
left temporal pole		9.27	-46	14	-36
left middle temporal gyrus		9	-56	2	-28
right precuneus	3605	9.41	4	-56	32
left precuneus		9.03	-4	-60	22
left precuneus		8.81	-12	-52	36
left superior medial frontal gyrus	4278	8.06	-10	48	36
left superior frontal gyrus		7.93	-12	48	22
right anterior cingulate cortex		6.81	8	54	10
left parahippocampal gyrus	203	6.7	-26	-38	-10
left lingual gyrus		4.63	-22	-46	-6
left lingual gyrus		3.94	-24	-54	-4
left anterior cingulate cortex	141	6.27	-2	10	-10
left thalamus	117	5.64	-4	-24	2
right thalamus		4.74	6	-26	0
right inferior frontal gyrus	248	4.89	46	34	-12
right inferior frontal gyrus		4.54	56	26	8
right inferior frontal gyrus		4.51	46	22	14
left cerebellum	122	4.83	-26	-74	-36
left cerebellum		4.02	-18	-80	-38
left cerebellum		3.63	-16	-70	-32

## **The role of mentalizing in the establishment of conceptual pacts**

Based on the results of our previous study, we were especially interested in studying how neural representations of conceptual pacts change as they are established and whether this would lead to increased or decreased engagement of the mentalizing network. The results of the ROI analysis revealed interesting changes in the recruitment of mentalizing areas over the course of repeated references in the communicative blocks. That is, the mentalizing network was recruited most strongly in the first round of the communicative blocks. Parameter estimates significantly decreased in the second round. Brain regions involved in mentalizing thus appear to play an especially important role during the initial coordination process.

We hypothesized that the difference between the communicative and individual blocks may be maximal in mentalizing areas following a misunderstanding. If the matcher did not accept or understand the director's initial proposal for a linguistic label, the director received feedback from the matcher and tried to describe the same figure again. We hypothesized that directors may need to engage in (additional) mentalizing to adjust their initial response based on the matcher's feedback in these situations. In line with this hypothesis, the ROI analysis revealed increased activity in the left TPJ, right TPJ and precuneus for the comparison between the repetitions and the first attempts, showing that repetitions indeed engage brain regions involved in mentalizing more than first attempts. This effect may also drive the preferential engagement of mentalizing areas during the first round of the communicative task, as this round contained the most attempts (see Figure 5.2B).

## **Whole-brain results**

The whole-brain results show that the neural differences between the communicative and individual blocks are most pronounced during the initial stages of establishing conceptual pacts. The directed interaction revealed large cortical and subcortical clusters. Interestingly, many of these clusters were bilateral. These findings suggest that the neural differences between the communicative and the individual blocks were mostly driven by the initial coordination process in the communicative blocks. Typically, participants settled on a set of linguistic labels during this round and reused these labels in subsequent rounds. Contrary to our expectations, the whole-brain results did not reveal activations in the key regions of the mentalizing network, with the exception of the right TPJ.

A closer look at the results of the direct comparison between the first rounds of the communicative and individual tasks suggests that at least some of the differences between the communicative and individual tasks are related to memory and attention processes. The parahippocampal gyrus and hippocampus are known to play an important role in the binding of pieces of contextual information in episodic memory (Aminoff, Kveraga, & Bar,

2013; Gabrieli, 1998; Squire, 1992), and the inferior parietal cortex supports the attentional influence on memory (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008). A possible explanation for at least part of the observed differences may thus be that the communicative task more strongly engaged areas involved in episodic memory encoding. This could be due to the need to encode more relevant context information in the communicative context (e.g., the communicative partner). Assuming that the neural differences are due to memory processes, it is interesting to note the relative absence of frontal clusters in most of the contrasts.

The neural differences in structures involved in episodic memory processes suggest that establishing conceptual pacts may at least partly engage domain-general episodic memory processes, as proposed by Horton & Gerrig (2005). This possibility may seem difficult to reconcile with the research showing successful encoding of conceptual pacts in patients with hippocampal amnesia, although a later study showed that the patients are impaired when tangram figures are highly visually similar (Duff et al., 2012). Future research will have to further clarify the link between memory encoding and retrieval in communicative and collaborative contexts and investigate the effect of the richness of the interactive context and the type of stimuli on these processes.

### Comparing encoding and retrieval

It is interesting to compare the results of the current study to the findings of our previous experiment (chapter 4). In that study, we found increased activity in the mPFC, the right TPJ, and the precuneus during the retrieval of collaboratively as compared to individually learned linguistic labels. However, in the current study, the neural representations of collaboratively and individually learned labels appear to become more similar over the course of the tasks, and the recruitment of brain regions involved in mentalizing decreases over the course of repeated references.

There are a number of possible explanations for the puzzling differences between the results of these studies. One possible explanation is that the encoding and retrieval of jointly established linguistic labels rely on distinct neural resources, or at least engage regions involved in mentalizing to a different extent. Alternatively, and perhaps more likely, the differences between the studies may be due to differences in the experimental set-up. While our previous study allowed for relatively free-flowing interaction during the behavioral session, the current experiment greatly restricted participants' interactive means. Although the current paradigm allowed for bidirectional communication, the simplified feedback and lack of partner-specific signals (e.g., voice) may have reduced the partner-specificity of the representations associated with the linguistic labels.

## **Conclusion**

In this study, we investigated how the process of coordinating meaning shapes the neural representations of conceptual pacts. Brain regions involved in mentalizing appear to be most strongly recruited during the initial stages of establishing conceptual pacts and following misunderstandings in a communicative context. The whole-brain results revealed neural differences in parietal, occipital and temporal brain regions between the communicative and the individual task. Similar to the ROI results, these neural differences were most pronounced during the initial stages of establishing linguistic labels, and following misunderstandings between the participants in the communicative task. In addition, we found evidence that the establishment of conceptual pacts at least partly relies on domain-general episodic memory processes.





General discussion





## General discussion

In this thesis, I have investigated how the communicative context in which we speak affects the cognitive and neural processes underlying language production. I focused especially on how speakers take into account and establish common ground to achieve mutual understanding. The experiments described in the second and third chapters focused on the way in which speakers take moment-by-moment changes in common ground into account. In the fourth and fifth chapters, I investigated how speakers build up common ground over the course of a series of interactions by establishing conceptual pacts. In this discussion chapter, I will first summarize the main findings of these experiments, followed by a discussion of the implications of these findings for psycholinguistic theories and possible avenues for future research.

### Summary of the findings

In chapter two, I used eye-tracking to investigate when and how common and privileged ground information affect utterance planning. Participants played a computerized version of a referential communication game in which a director described objects in an array for a matcher. We manipulated common ground by occluding certain objects from the matcher's point of view. On critical trials, the director had to adjust their description based on this perspective difference to unambiguously describe the target objects to the matcher. Speakers mainly, but not always, produced referring expressions that took into account their addressee's visual perspective. Eye-tracking data revealed that speakers distinguished between common and privileged ground from the earliest stages of utterance planning, but did not completely ignore objects in privileged ground. Finally, we did not find evidence that taking common ground into account requires additional planning time.

In chapter three, I adapted the paradigm from the second chapter to study the neural processes that allow us to take moment-to-moment changes in common ground into account by means of fMRI. In addition to comparing situations in which the distinction between common and privileged ground information was relevant to utterance planning, I directly compared communicative and non-communicative planning. I found that the medial prefrontal cortex, a core region of the mentalizing network, is especially sensitive to communicative contexts in which speakers have to take their addressee's needs into account in order to communicate efficiently. In addition, I found neural differences between the communicative and the non-communicative settings before speakers started to plan their utterances, suggesting that they continuously keep track and update common ground in a communicative context.

The fourth chapter looked at how jointly establishing and learning linguistic labels in a communicative context affects the memory networks involved during the retrieval of the labels. In this experiment, participants learned labels for abstract figures in three conditions: they generated labels in a communication game with another participant, they came up with labels by themselves, and they learned a set of unrelated labels. Participants then retrieved all labels in the MRI scanner during a communication task. We found that the retrieval of collaboratively generated labels as compared to individually learned labels engages brain regions involved in mentalizing, including the medial prefrontal cortex, the right temporoparietal junction and the precuneus. These findings show that the process of establishing linguistic labels in a communicative context can affect the neural networks involved in the retrieval of these labels.

In chapter five, I investigated how the process of coordinating meaning shapes the neural representations of jointly established linguistic labels (conceptual pacts). Participants in this fMRI experiment repeatedly described abstract figures in either a communicative context or an individual context. We found that mentalizing areas are most strongly engaged during the initial stages of establishing conceptual pacts with a communication partner and following misunderstandings between participants in the communicative context. In addition to the neural differences in mentalizing areas, brain regions involved in episodic memory encoding were more strongly engaged in the initial stages of establishing linguistic labels in the communicative task as compared to the individual task.

### **The role of common ground in speech planning**

The second and third chapters focused on how and when the distinction between common and privileged ground affects language production. Behaviorally, we did not find a significant main effect of condition on planning duration in either of these studies. Null findings are hard to interpret, especially given the modest sample sizes in these studies, but these results do suggest that taking another person's visual perspective into account in a communicative context can be a relatively fast and efficient process. The results of the fMRI study in chapter three suggest that speakers' fast adaptation to their addressee is achieved as a result of the recruitment of additional neural resources when speakers need to take common ground into account during utterance planning. Brain regions involved in mentalizing and cognitive control play an important role in this process.

The eye-tracking results from the second chapter showed that speakers distinguish between common and privileged ground from the earliest stages of utterance planning. Despite this early distinction, these data also showed that speakers could not fully ignore privileged ground information. Combined, these findings suggest that common ground exerts an early, if incomplete effect on utterance planning. A similar picture emerged from the

speakers' responses. While speakers mainly produced referring expressions that took into account their addressee's visual perspective, they occasionally failed to do so. This did not only happen when they risked being overinformative, but even when the addressee would not be able to infer the intended referent. These findings fit well with Constraint-Based Processing models that predict an early and probabilistic influence of common ground on language processing (Brennan & Hanna, 2009; Hanna, Tanenhaus & Trueswell, 2003; Horton & Gerrig, 2002; Tanenhaus & Trueswell, 1995). In this view, common ground is one of multiple constraints (e.g., saliency, linguistic context) that can guide language processing.

### **Common ground, mentalizing and memory**

In the fourth and fifth chapters, I turned my attention to the effects of communication on memory representations. Using fMRI, I found that the retrieval of conceptual pacts as compared to individually generated labels recruits brain regions involved in mentalizing. These findings show that the process of grounding referring expressions with someone else can lead to at least partly neurally distinguishable representations at retrieval. In a follow-up experiment, I aimed to clarify the role of mentalizing in the encoding and retrieval of conceptual pacts by investigating how the process of coordinating meaning shapes the neural representations of jointly established linguistic labels. The results of this study showed that brain regions involved in mentalizing are most strongly recruited during the initial stages of establishing conceptual pacts and following feedback from the addressee. Most differences between the communicative and the individual contexts disappeared over the course of repeated references.

What do these findings tell us about the memory representations of conceptual pacts? The results of chapters four and five point to the involvement of the mentalizing network in both the encoding and retrieval of conceptual pacts. The decreasing involvement of brain regions involved in mentalizing in chapter five was unexpected based on the results of chapter four. One possible explanation is that the free-flowing, interactive nature of the encoding task used in chapter four resulted in richer and stronger partner-specific memory traces than the simplified communication paradigm used in chapter five. This possibility raises interesting questions and methodological considerations for future research on the effects of communication and collaboration on memory. A major challenge will be to create naturalistic and multimodal communicative contexts in the lab.

The results of chapter five also showed that episodic memory processes appear to play a role in the encoding and retrieval of conceptual pacts. We found increased activity in the hippocampus and parahippocampal gyrus during the initial stages of creating linguistic labels in a communicative as compared to an individual context. This fits well with Horton & Gerrig's proposal (2005) that conversation partners create rich episodic memory traces of

their interactions, thus linking interlocutors and conceptual pacts in memory. The presence of a conversational partner can then serve as a cue that reactivates overlapping information in memory, such as previously established conceptual pacts. Future research will further have to clarify the contribution of mentalizing and episodic memory processes and their interplay in the encoding and retrieval of conceptual pacts.

### **Implications for the neurobiology of language**

This thesis illustrates that research on the neurobiology of language should not ignore the social and communicative context in which language is used. The results show that the communicative context in which we speak influences the cognitive and neural mechanisms underlying language production. Neurobiological theories of language processing should therefore account for the role of social and communicative factors. This will require additional multidisciplinary research that links the neurobiology of language to fields such as social cognition, memory, action and attention, as well as to the behavioral psycholinguistic literature on this topic. In addition, this approach will require the development of novel paradigms and methods to study language processing in interactive settings (see the section on avenues for future research below).

The results described in this thesis do not point to one specific brain region or network that is systematically recruited when we use language to communicate with others. While brain regions involved in mentalizing and theory of mind appear to play an important role, I also found neural differences in areas involved in cognitive control, visual attention and episodic memory. These results suggest that communicative language processing may be highly dependent on the interaction between the classical language network and other cognitive systems. The precise interactions between these systems probably depend on the task or situation at hand. Future research will further have to elucidate under what conditions and how these systems interact.

### **Avenues for future research**

The research described in this thesis raises a number of questions for further research. Below I highlight some interesting avenues for future research related to the generalizability of the findings, the role of individual differences, and developmental changes in children's sensitivity to common ground in language processing.

One important question is to what extent the findings described in the previous chapters generalize to more realistic conversational settings. The paradigms I have used in my experiments all featured relatively simple two-person communicative interactions. In real-life settings, conversations often involve more than two conversational partners, which may make keeping track of the common ground shared with all partners considerably more

complicated. Similarly, the referential communication task used in the second and third chapters made the perspective difference between speakers and listeners visually salient. Taking common ground into account in this situation may therefore involve relatively low-level visual-perspective taking. However, in real-life conversations, speakers sometimes have to infer what their interlocutor believes or feels in addition to what they can see, which may require different perspective-taking processes.

A major challenge for future studies on communicative language processing will be to develop more naturalistic paradigms that tap into the complexity and multimodal nature of real-life interactions. Virtual reality appears to be an especially promising method in this respect, as it offers the possibility to develop experimental paradigms that are rich, controlled and compatible with neuroimaging. In addition, simultaneously measuring neural activity in several interacting participants (i.e. hyperscanning) may provide interesting new insights in how mutual understanding emerges. For example, one could study how and where brain-to-brain coupling increases as interlocutors establish conceptual pacts or build up common ground over the course of repeated interactions. Such novel methods will allow us to test whether current findings extend to more realistic and complex environments, and may capture additional factors that can influence communicative language processing.

The research described in this thesis points to an important role for the theory of mind or mentalizing network in communicative language production, but future research will have to further specify the exact contribution of this network. Many authors have argued that the brain regions that make up the mentalizing network subserves particular sub-processes of mentalizing (e.g., Amodio & Frith, 2006; Mitchell et al., 2006; Saxe & Kanwisher, 2003). Yet, the exact function of these different sub-regions remains a hotly debated issue in the social cognition literature. For example, while some have argued that the right temporoparietal junction selectively subserves the attribution of beliefs to other people (e.g., Saxe & Kanwisher, 2003; Saxe & Powell, 2006), others have proposed that this region plays a more general role in reorienting attention (e.g., Mitchell, 2008). The experiments described in this thesis were not designed to tease apart the functions of the sub-regions of the mentalizing network, but my results shed at least some light on their role. For example, in chapter 3 I found that the medial prefrontal cortex is especially sensitive to communicative contexts in which speakers have to take their addressee's needs into account in order to communicate efficiently. New developments in the social cognition literature will undoubtedly lead to new hypotheses about the role of the sub-regions of the mentalizing network in communicative language production.

The results reported in this thesis also suggest that a number of other domain-general cognitive functions, including attention, control and episodic memory processes contribute to speakers' ability to consider and establish common ground. One approach to further specify the contributions of such domain-general cognitive functions is to study individual differences. So far, studies of individual differences using referential communication tasks have found effects of mentalizing (Sidera, Perpiñà, Serrano, & Rostan, 2016), inhibitory control (Brown-Schmidt, 2009; Nilsen, Varghese, Xu, & Fecica, 2015; Nilsen & Graham, 2009; Wardlow, 2013) and working memory abilities (Nilsen et al., 2015; Wardlow, 2013).

Another promising approach is to study the development of children's sensitivity to common ground. Given the relatively slow development of children's theory of mind skills during the preschool years, studying their sensitivity to common ground at different ages can provide interesting insights into the role and nature of perspective-taking in this process. The visual perspective-taking skills needed to infer which objects a person with a different visual perspective can or cannot see develop by 24 months (Masangkay et al., 1974; Moll & Tomasello, 2006). It is therefore perhaps unsurprising that 5- to 6-year-olds are sensitive to the distinction between common and privileged ground information in both language production and comprehension, as tested using a simple paradigm similar to the task used in chapters two and three (Nadig & Sedivy, 2002). It remains to be seen whether younger children show similar sensitivity. Perhaps more interestingly, four- and six-year-old children can establish conceptual pacts with their peers (Köymen, Schmerse, Lieven, & Tomasello, 2014), and even three-year-old children do not expect their conversation partner to break a jointly established conceptual pact (Matthews, Lieven, & Tomasello, 2010). Given that children only begin to pass false belief tasks around the age of four (Wellman, Cross, & Watson, 2001), these findings suggest that high-level mental reasoning may not be a precondition for the emergence of partner-specific effects.

## Conclusion

In this thesis, I have investigated the role of common ground in communicative language production by combining existing psycholinguistic paradigms with eye-tracking and fMRI. The results of the experiments show that communicative factors such as common ground can affect both online linguistic planning (chapters 2 and 3), as well as the encoding and retrieval of jointly established linguistic labels over the course of communicative interactions (chapters 4 and 5). These findings suggest that a complete theory of the neurobiology of language needs to account for social and communicative influences on language processing.





Appendix





## References

- Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, 17(8), 379–390.
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews. Neuroscience*, 7(4), 268–77.
- Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Barber, S. J., Rajaram, S., & Aron, A. (2010). When two is too many: Collaborative encoding impairs memory. *Memory & Cognition*, 38(3), 255–264.
- Barber, S. J., Rajaram, S., & Paneerselvam, B. (2012). The collaborative encoding deficit is attenuated with specific warnings. *Journal of Cognitive Psychology*, 24(8), 929–941.
- Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, 109(1), 18–40.
- Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1176.
- Bates, D., Mächler, M., Bolker, B.M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, 19(12), 2767–2796.
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott international*, 5(9/10), 341–345.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 22(6), 1482–93.
- Brennan, S. E., Galati, A., & Kuhlén, A. K. (2010). Two Minds, One Dialog: Coordinating Speaking and Understanding. In B. H. Ross (Ed.), *Psychology of Learning and Motivation, Vol. 53* (pp. 301–344). Burlington: Academic Press/Elsevier.
- Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science*, 1(2), 274–291.
- Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage*, 16(2), S497.
- Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, 16(5), 893–900.
- Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialog and Discourse*, 2(1), 11–33.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscribed conversation: A targeted language game approach. *Cognitive science*, 32(4), 643–684.
- Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews. Neuroscience*, 9(8), 613–25.

- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563.
- Caruana, N., Brock, J., & Woolgar, A. (2015). A frontotemporoparietal network common to initiating and responding to joint attention bids. *NeuroImage*, 108, 34–46.
- Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., & Walter, H. (2007). The intentional network: How the brain reads varieties of intentions. *Neuropsychologia*, 45(13), 3105–3113.
- Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G., & Walter, H. (2014). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Social Cognitive and Affective Neuroscience*, 9(7), 909–916.
- Clark, H.H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Clark, H.H., & Brennan, S. E. (1991). Grounding in Communication. In L.B. Resnick, J.M. Levine & S.D. Teasley (Eds.), *Perspectives on Socially Shared Cognition*, 127–149. Washington: APA Books.
- Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, 58(2), 332-373.
- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1), 62-81.
- Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical Issues In Natural Language Processing*, Vol. 2. (pp. 57-63). New York: Association for Computing Machinery.
- Clark, H.H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A.K. Joshi, B. Webber & I. Sag, (Eds.), *Elements of Discourse Understanding* (pp. 10–63). Cambridge: Cambridge University Press.
- Clark, H. H., & Murphy, G. L. (1983). Audience design in meaning and reference. In J. F. LeNy & W. Kintsch (Eds.), *Language and comprehension* (pp. 287-299). Amsterdam: North-Holland Publishing Co.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1-39.
- Cleret de Langavant, L., Remy, P., Trinkler, I., McIntyre, J., Dupoux, E., Berthoz, A., & Bachoud-Lévi, A. C. (2011). Behavioral and neural correlates of communication via pointing. *PLoS ONE*, 6(3), 1–10.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Corbetta, M., & Shulman, G. L. (2002). Control of Goal-Directed and Stimulus-Driven Attention in the Brain. *Nature Reviews Neuroscience*, 3(3), 215–229.
- 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.
- Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The Use of Definite References Signals Declarative Memory Evidence From Patients With Hippocampal Amnesia. *Psychological Science*, 22(5), 666-673.

- Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, 9(1), 140–6.
- Duff, M. C., Warren, D. E., Gupta, R., Vidal, J. P. B., Tranel, D., & Cohen, N. J. (2012). Teasing apart tangrams: Testing hippocampal pattern separation with a collaborative referencing paradigm. *Hippocampus*, 22, 1087–1091.
- Dumontheil, I., Hillebrandt, H., Apperly, I. A., & Blakemore, S.-J. (2012). Developmental differences in the control of action selection by social information. *Journal of Cognitive Neuroscience*, 24(10), 2080–95.
- Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S.-J. (2010). Taking perspective into account in a communicative task. *NeuroImage*, 52(4), 1574–83.
- Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54(4), 554–573.
- Ethofer, T., Gschwind, M., & Vuilleumier, P. (2011). Processing social aspects of human gaze: A combined fMRI-DTI study. *NeuroImage*, 55(1), 411–419.
- Farthing, J. P., Cummine, J., Borowsky, R., Chilibeck, P. D., Binsted, G., & Sarty, G. E. (2007). False activation in the brain ventricles related to task-correlated breathing in fMRI speech and motor paradigms. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 20(3), 157–168.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–534.
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63, 287–313.
- 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 B: Biological Sciences*, 365(1537), 165–176.
- Gabrieli, J. D. E. (1998). Cognitive Neuroscience of Human Memory. *Annual Review Psychology*, 49(1), 87–115.
- Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, 62(1), 35–51.
- Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech Arts* (pp. 41–58). New York: Academic Press.
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49(1), 43–61.
- Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, 108(3), 831–836.
- Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, 22(7), 1114–1139.
- Horton, W. S., & Brennan, S. E. (2016). The Role of Metarepresentation in the Production and Resolution of Referring Expressions. *Frontiers in Psychology*, 7(July), 1–12.

- Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language, 47*(4), 589–606.
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition, 96*(2), 127–42.
- Horton, W.S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91-117.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal, 50*(3), 346-363.
- Huetig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*(2), 151–171.
- Kampe, K. K. W., Frith, C. D., & Frith, U. (2003). "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *The Journal of Neuroscience, 23*(12), 5258–63.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32–38.
- Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science, 7*(2), 46-50.
- Köymen, B., Schmerse, D., Lieven, E., & Tomasello, M. (2014). Young children create partner-specific referential pacts with peers. *Developmental Psychology, 50*(10), 2334–2342.
- Kuhlen, A. K., Allefeld, C., Anders, S., & Haynes, J. D. (2015). Towards a multi-brain perspective on communication in dialogue. In R. Willems (Ed.), *Cognitive Neuroscience of Natural Language Use* (pp. 182-200). Cambridge University Press.
- Kuhlen, A. K., & Brennan, S. E. (2012). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin & Review, 20* (1), 54-72.
- Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development, 45*, 357-366.
- Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology, 46*(4), 749–60.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*(2), 201–213.
- Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex, 18*(2), 262-271.
- Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences, 99*(23), 15238-15243.
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding-Specific Effects of Social Cognition on the Neural Correlates of Subsequent Memory. *Journal of Neuroscience, 24*(21), 4912–4917.

- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*(4), 655–63.
- Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*(3), 603–613.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336.
- Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*(2), 220–49.
- Nilsen, E. S., Varghese, A., Xu, Z., & Fecica, A. (2015). Cognitive Development Children with stronger executive functioning and fewer ADHD traits produce more effective referential statements. *Cognitive Development*, *36*, 68–82.
- 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), 1–13.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, *308*, 214–216.
- Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 781–795.
- 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.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816–847.
- Prince, S. E., Daselaar, S. M., & Cabeza, R. (2005). Neural Correlates of Relational Memory: Successful Encoding and Retrieval of Semantic and Perceptual Associations. *Journal of Neuroscience*, *25*(5), 1203–1210.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Read, R.C. (1965). Tangrams: 330 Puzzles. New York: Dover Publications.
- Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *NeuroImage*, *50*(4), 1639–47.
- Redcay, E., Ludlum, R. S., Velnoskey, K. R., & Simren, K. (2015). Communicative Signals Promote Object Recognition Memory and Modulate the Right Posterior STS. *Journal of Cognitive Neuroscience*, 1–12.
- 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.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination. *Psychological Science*, *18*(5), 407–413.

- Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, *56*(2), 129–140.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*, 696–735.
- Sassa, Y., Sugiura, M., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2007). Cortical mechanism of communicative speech production. *NeuroImage*, *37*(3), 985–92.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporoparietal junction in “theory of mind.” *NeuroImage*, *19*(4), 1835–1842.
- Saxe, R., & Powell, L. J. (2006). It’s the thought that counts specific brain regions for one component of theory of mind. *Psychological science*, *17*(8), 692–699.
- Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., . . . Vogeley, K. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, *22*(12), 2702–2715.
- Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1–24.
- 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*(November), 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 and Biobehavioral Reviews*, *42*, 9–34.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.
- Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development*, *38*, 1–38.
- Sidera, F., Perpiñà, G., Serrano, J., & Rostan, C. (2016). Why Is Theory of Mind Important for Referential Communication? *Current Psychology*, 1-16.
- Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(43), E4687–E4696.
- Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research*, *17*(1), 75–82.
- Spiegelhalder, K., Ohlendorf, S., Regen, W., Feige, B., Tebartz van Elst, L., Weiller, C., Hennig, J., Berger, M., & Tüscher, O. (2014). Interindividual synchronization of brain activity during live verbal communication. *Behavioural Brain Research*, *258*, 75–79.
- Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological review*, *99*(2), 195–231.
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14425–30.

- Stolk, A., Noordzij, M. L., Volman, I., Verhagen, L., Overeem, S., van Elswijk, G., ... Toni, I. (2014). Understanding communicative actions: a repetitive TMS study. *Cortex*, *51*, 25–34.
- Stolk, A., Verhagen, L., Schoffelen, J.-M., Oostenveld, R., Blokpoel, M., Hagoort, P., van Rooij, I., & Toni, I. (2013). Neural mechanisms of communicative innovation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14574–14579.
- Tanenhaus, M., & Trueswell, J. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception: Vol. 11. Speech and Language* (pp. 217–262). San Diego, CA: Academic Press.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*(5), 352.
- Van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S.-A. (2012). Flexibility in embodied lexical-semantic representations. *Journal of Cognitive Neuroscience*, *24*(10), 2237–2247.
- Vanlangendonck, F., Willems, R. M., Menenti, L., & Hagoort, P. (2016). An early influence of common ground during speech planning. *Language, Cognition and Neuroscience*, *31*(6), 741–750.
- Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, *30*(3), 829–88.
- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564–584.
- Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, *16*(10), 1854–1863.
- Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic bulletin & review*, *20*(4), 766–772.
- Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1466–1481.
- Wardlow Lane, L., Groisman, M., Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science*, *17*(4), 273–277.
- Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: collaborative and individual processes in remembering. *Journal of experimental psychology: Learning, memory, and cognition*, *23*(5), 1160–1175.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child development*, *72*(3), 655–684.
- Willems, R. M., de Boer, M., de Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychological Science*, *21*(1), 8–14.
- Yoon, S. O., Koh, S., & Brown-Schmidt. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, *19*(4), 699–707.



## Nederlandse samenvatting

### Communicatie en common ground

Je gebruikt taal in de eerste plaats om met anderen te communiceren. Of je nu roddelt, een drankje bestelt of lesgeeft, meestal spreek je met *iemand* en niet tegen jezelf. Het feit dat je taal gebruikt om te communiceren heeft gevolgen voor de manier waarop je je uitdrukt. Als je met je driejarige neefje spreekt, gebruik je waarschijnlijk simpelere woorden dan wanneer je met een collega overlegt. Of beeld je in dat je een verdwaalde vriend aan de telefoon moet uitleggen hoe hij je huis kan vinden. Om duidelijke aanwijzingen te kunnen geven, is het handig om je in zijn schoenen te verplaatsen. Uit deze voorbeelden blijkt dat je taalgebruik wordt beïnvloed door de communicatieve context waarin je spreekt.

Hoe zorgt je brein ervoor dat je je taalgebruik schijnbaar moeiteloos kunt aanpassen aan je gesprekspartner? Er is hierover nog vrij weinig geweten, omdat psycholinguïstisch onderzoek doorgaans geen rekening houdt met communicatieve en sociale invloeden op taalgebruik. In dit proefschrift heb ik daarom onderzocht hoe de communicatieve context waarin je spreekt van invloed is op de manier waarop je spreekt en de onderliggende hersenprocessen.

De rode draad in dit proefschrift is het concept **common ground**, letterlijk te vertalen als "gedeelde grond". *Common ground* verwijst naar de overtuigingen en kennis die gesprekspartners delen. Deze gedeelde informatie speelt een belangrijke rol in conversaties. Stel je voor dat ik tegen je zeg: "Ik zag Jan gisteren bij de Albert Heijn." Jij begrijpt dan alleen wie ik gisteren heb gezien als we allebei weten wie Jan is en we van elkaar weten dat we Jan kennen. In dit geval maak ik gebruik van onze *common ground* ("wij kennen allebei Jan") om een duidelijke zin te formuleren.

*Common ground* kan je ook opbouwen tijdens een conversatie. Misschien hebben wij bijvoorbeeld meerdere vrienden die Jan heten en verloopt de conversatie als volgt:

- "Ik zag Jan gisteren bij de Albert Heijn."
- "Jan van de Franse les?"
- "Nee, Jan van Sofie"
- "Hoe gaat het met hem?"

Met enkele zinnen hebben we samen bepaald welke Jan we bedoelen en een afspraak gemaakt over hoe we naar hem kunnen verwijzen. We kunnen het daardoor tijdens de rest van ons gesprek over "Jan van Sofie" hebben zonder dat dit tot verwarring leidt. Met dit soort afspraak, ook een **conceptueel pact** genoemd, zorgen we ervoor dat de verwijzing "Jan van Sofie" deel uitmaakt van onze *common ground*.

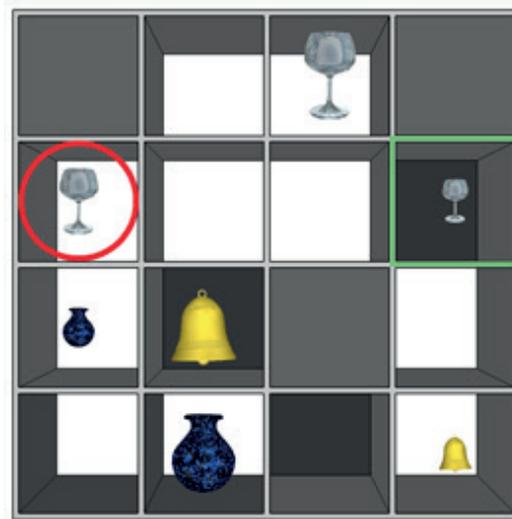
In dit proefschrift rapporteer ik de resultaten van vier experimenten waarin ik heb onderzocht hoe *common ground* ons taalgebruik en de onderliggende hersenprocessen beïnvloedt. In hoofdstukken twee en drie heb ik vooral gekeken naar wanneer en hoe *common ground* taalproductie beïnvloedt. In hoofdstukken vier en vijf heb ik de relatie tussen communicatie en geheugen onderzocht, waarbij ik heb bestudeerd hoe je conceptuele pactsen opbouwt en onthoudt.

## Common ground en taalproductie

In **hoofdstuk twee** heb ik onderzocht wanneer en hoe je tijdens het spreken rekening houdt met welke informatie je deelt met je gesprekspartner (*common ground*) en welke informatie je niet deelt (*privileged ground*). Er bestaan twee theorieën over wanneer en hoe mensen rekening houden met *common ground* terwijl ze plannen wat ze willen zeggen. Volgens de ene theorie houden mensen tijdens het plannen al vroeg rekening met welke informatie ze delen met hun gesprekspartner. Volgens de andere theorie plannen mensen wat ze willen zeggen aanvankelijk vanuit hun eigen perspectief en passen dit pas later aan als blijkt dat dit nodig is.

Ik heb een interactieve taak ontwikkeld om te testen of sprekers vroeg of laat rekening houden met *common ground*. Aan dit experiment namen telkens twee proefpersonen deel, een spreker en een luisteraar. De spreker en de luisteraar zagen tegenovergestelde zijden van een kast die was ingedeeld in vakjes met objecten. Sommige vakjes met objecten waren open aan beide zijden en dus zichtbaar voor spreker en luisteraar; andere vakjes met objecten waren maar aan één zijde open en dus enkel zichtbaar voor de spreker of voor de luisteraar. Door vakjes te openen of te sluiten kon ik manipuleren welke objecten zichtbaar waren voor beide proefpersonen (gedeelde informatie of *common ground*) en welke objecten alleen zichtbaar waren voor een van de proefpersonen (niet-gedeelde informatie of *privileged ground*).

De spreker kreeg telkens de opdracht om een van de objecten in de kast te beschrijven voor de luisteraar. De luisteraar probeerde dan op dit object te klikken. Soms zag de spreker extra objecten die de luisteraar niet kon zien en moest de spreker hiermee rekening houden om efficiënt met de luisteraar te communiceren. Als je in de kast in figuur A.1 bijvoorbeeld het middelgrote glas (aangeduid met de rode cirkel) beschrijft voor iemand die aan de andere kant van de kast zit, moet je er rekening mee houden dat deze persoon het kleinste glas (hier aangeduid met het groene vierkant) niet kan zien. Als je rekening houdt met wat de ander kan zien, zou je het glas omschrijven als "het kleine glas". Als je dat niet doet, zou je het omschrijven als "het middelgrote glas", en weet de andere persoon niet welk glas je bedoelt. Hij of zij ziet namelijk slechts twee glazen.



**Figuur A.1:** Voorbeeld van de objecten die de proefpersonen in de taken uit hoofdstukken 2 en 3 beschreven.

Op basis van de omschrijvingen van de sprekers kon ik vaststellen of sprekers rekening hielden met welke objecten hun gesprekspartner kon zien: zeiden ze “het kleine glas” of “het middelgrote glas”? Ik vond dat sprekers in hun antwoorden meestal rekening hielden met wat de luisteraars konden zien. Om te onderzoeken wanneer sprekers bij het plannen van hun omschrijvingen rekening hielden met *common ground*, heb ik ook hun oogbewegingen gemeten met een eye-tracker. Zodoende kon ik precies meten hoe lang en wanneer de sprekers naar objecten keken die zichtbaar of onzichtbaar waren voor de luisteraar. Hieruit bleek dat sprekers al vroeg tijdens het plannen van wat ze willen zeggen minder kijken naar objecten die niet zichtbaar zijn voor de luisteraars. De resultaten van dit experiment bevestigen dus de eerste hypothese dat sprekers tijdens het plannen al vroeg rekening houden met *common ground*.

In **hoofdstuk drie** heb ik onderzocht welke hersengebieden actief zijn wanneer je rekening houdt met *common ground*. Uit eerder onderzoek is gebleken dat extra hersengebieden actief worden wanneer mensen iets beschrijven met als doel om met iemand anders te communiceren dan wanneer ze niet het doel hebben om te communiceren. De bijkomende hersengebieden die daarbij actief zijn, zijn ook actief wanneer je je mentaal in de schoenen van iemand anders probeert te verplaatsen. Dit is bijvoorbeeld het geval wanneer je moet begrijpen dat iemand anders een situatie anders interpreteert dan jij omdat ze over andere informatie beschikken. De hersengebieden die actief worden wanneer je je mentaal in de schoenen van een ander probeert te verplaatsen (de mediale frontale cortex, de temporoparietale juncties en de precuneus) worden samen ook het *mentalizing network*

genoemd. In dit experiment wilde ik testen of het *mentalizing network* vooral actief is wanneer je rekening moet houden met *common ground*.

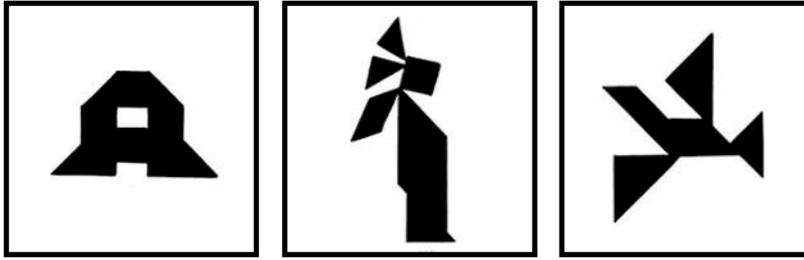
Ook aan dit experiment deden telkens twee proefpersonen mee. De ene proefpersoon lag in de MRI-scanner en de andere zat aan een computer buiten het MRI-lab. Tijdens het experiment voerde de proefpersoon in de MRI-scanner afwisselend een communicatieve taak met de andere proefpersoon en een taak alleen uit. De communicatieve taak met de andere proefpersoon was gebaseerd op de taak uit hoofdstuk 2: de proefpersoon in de MRI-scanner beschreef objecten in een kast met open en gesloten vakjes en de andere proefpersoon probeerde op basis van deze beschrijvingen op de objecten te klikken. Soms moest de proefpersoon in de scanner daarvoor rekening houden met *common ground* (bijv. wanneer hij/zij middelgrote glas in figuur A.1 moet beschrijven) en soms niet (bijv. als hij/zij de kleine vaas moet beschrijven). De individuele taak leek sterk op de communicatieve taak, behalve dat de proefpersoon buiten de MRI-scanner niet meedeed. De proefpersoon in de scanner moest in deze taak dus geen rekening houden met welke objecten zichtbaar waren voor de andere proefpersoon. Tijdens beide taken heb ik door middel van fMRI de hersenactiviteit van de proefpersonen in de MRI-scanner gemeten.

Met deze opzet kon ik twee interessante vergelijkingen onderzoeken: de vergelijking tussen taalproductie in een communicatieve en een niet-communicatieve context, en de vergelijking binnen de communicatieve taak tussen communicatieve situaties waarin je rekening moet houden met *common ground* en situaties waarin dat niet nodig is. Voor de vergelijking tussen spreken in een communicatieve en een niet-communicatieve context vond ik, in tegenstelling tot vorige studies, slechts een klein neuraal verschil. Ik vond echter grote verschillen toen ik binnen de communicatieve taak vergeleek welke gebieden actief zijn voor de vergelijking tussen communicatieve situaties waarin je rekening moet houden met *common ground* en situaties waarin dat niet nodig is. Op het moment dat je rekening moet houden met wat je gesprekspartner kan zien, wordt een groot deel van het *mentalizing network* actief. Mijn resultaten zorgen dus voor een interessante aanvulling op eerder onderzoek. Vorige studies hebben gevonden dat het *mentalizing network* meer actief is wanneer je het doel hebt te communiceren met iemand anders dan wanneer je dat doel niet hebt. Mijn resultaten laten echter zien dat dit netwerk niet altijd even actief is in communicatieve situaties: het is voornamelijk actief wanneer je rekening moet houden met je gesprekspartner om efficiënt te communiceren.

## Common ground en geheugen

In het vierde en vijfde hoofdstuk heb ik de invloed van communicatie op het geheugen onderzocht. Ik heb bestudeerd hoe mensen conceptuele pactsen opbouwen en onthouden, en hoe conceptuele pactsen verschillen van beschrijvingen die je op je eentje leert. In **hoofdstuk vier** heb ik onderzocht of beschrijvingen die je met een ander hebt bedacht op een andere manier worden opgeslagen in je hersenen dan beschrijvingen die je alleen hebt bedacht. De aanleiding voor dit experiment was een studie over patiënten met amnesie. De onderzochte patiënten konden amper nieuwe verbanden leren als gevolg van hersenletsel in de hippocampus, het belangrijkste geheugencentrum in het brein. Opmerkelijk genoeg bleek dat deze patiënten in staat waren om in een communicatieve taak met een andere proefpersoon nieuwe beschrijvingen voor plaatjes te leren. Deze resultaten suggereren dat leren in een communicatieve context mogelijk deels via andere hersenprocessen verloopt dan leren in een niet-communicatieve, individuele context. Deze hypothese heb ik in het volgende experiment getest.

Aan dit experiment deden telkens twee proefpersonen mee en het experiment bestond uit twee delen. Tijdens het eerste deel leerde een van de proefpersonen in drie taken beschrijvingen voor abstracte plaatjes: een communicatieve taak, een individuele taak en een arbitraire taak. Figuur A.2 bevat een paar voorbeelden van deze plaatjes. In de communicatieve taak beschreef de proefpersoon plaatjes op een scherm zodat de andere proefpersoon op de beschreven plaatjes kon klikken. De proefpersonen konden hierbij vragen stellen en aanwijzingen geven aan elkaar, en bouwden zo samen beschrijvingen of conceptuele pactsen voor de plaatjes op. In de individuele taak bedacht de proefpersoon op zijn/haar eentje beschrijvingen voor een nieuwe set plaatjes. In de arbitraire taak, ten slotte, kreeg dezelfde proefpersoon een derde set abstracte plaatjes te zien met ongerelateerde beschrijvingen die hij/zij van buiten moest leren. Tijdens het tweede deel van het experiment lag deze proefpersoon in de MRI-scanner en kon ik zijn/haar hersenactiviteit meten. De proefpersoon kreeg alle plaatjes uit het eerste deel van het experiment opnieuw te zien en produceerde hardop de beschrijvingen die hij/zij hiervoor had geleerd. Ondertussen probeerde de andere proefpersoon op basis van deze beschrijvingen op de beschreven plaatjes te klikken.



**Figuur A.2:** Voorbeelden van de plaatjes waarvoor de proefpersonen in de experimenten uit hoofdstukken 4 en 5 beschrijvingen leerden.

Ik wilde met dit experiment onderzoeken of beschrijvingen die proefpersonen met iemand anders hebben bedacht in een communicatieve context anders worden opgeslagen in de hersenen dan beschrijvingen die ze alleen hebben bedacht. De gedragsdata lieten een interessant patroon zien: de omschrijvingen die proefpersonen samen hadden bedacht waren gemiddeld het langste, maar toch konden proefpersonen zich tijdens de taak in de MRI-scanner deze beschrijvingen het snelste en beste herinneren. Met de fMRI-resultaten kon ik onderzoeken welke hersengebieden actief waren terwijl proefpersonen zich de beschrijvingen probeerden te herinneren in de scanner. Ik was vooral geïnteresseerd in de vergelijking tussen het ophalen van beschrijvingen die proefpersonen met iemand anders hadden bedacht en beschrijvingen die ze zelf hadden bedacht. Ik vond dat het *mentalizing network* (de mediale frontale cortex, de temporoparietale juncties en de precuneus) meer actief was wanneer proefpersonen de samen bedachte beschrijvingen ophaalden dan wanneer ze de alleen bedachte beschrijvingen ophaalden. Dit is de eerste studie die aantoont dat het samen bedenken van beschrijvingen een blijvend effect heeft op de processen die betrokken zijn bij het later ophalen van de beschrijvingen uit je geheugen.

In **hoofdstuk 5** heb ik onderzocht hoe de neurale verschillen tussen samen en alleen bedachte beschrijvingen uit hoofdstuk 4 ontstaan. In dit experiment liet ik proefpersonen in de MRI-scanner beschrijvingen voor abstracte plaatjes bedenken in een communicatieve of een niet-communicatieve taak. Deze taken leken sterk op de communicatieve en de individuele taken uit hoofdstuk 4. In de communicatieve taak probeerde de proefpersoon in de MRI-scanner de plaatjes te beschrijven voor een proefpersoon buiten de scanner. De proefpersoon buiten de scanner probeerde op basis van deze beschrijvingen op het juiste plaatje te klikken. Als de beschrijving van de proefpersoon in de scanner niet helemaal duidelijk was, kon de proefpersoon buiten de scanner feedback geven aan de andere proefpersoon. Op deze manier bedachten de proefpersonen samen beschrijvingen voor de plaatjes. In de individuele taak bedacht de proefpersoon in de scanner individueel beschrijvingen voor een andere set plaatjes. In beide taken beschreef deze proefpersoon alle plaatjes meermaals.

Ik wilde met dit experiment onderzoeken hoe de neurale verschillen tussen samen en alleen bedachte beschrijvingen ontstaan en wanneer deze verschillen het grootst zijn. Is dat het geval wanneer je voor het eerst een beschrijving voor een plaatje bedenkt of pas na enkele herhalingen? Ik was hierbij vooral geïnteresseerd in verschillen binnen het *mentalizing network*, omdat ik daar grote verschillen vond in hoofdstuk 4. Binnen het *mentalizing network* bleek het verschil tussen het samen en alleen bedenken van beschrijvingen het grootst op het moment dat proefpersonen voor het eerst een beschrijving voor een plaatje bedachten of wanneer ze net feedback hadden kregen van hun gesprekspartner. Waarschijnlijk is dit het geval omdat ze vooral dan rekening houden met hun gesprekspartner om het plaatje zo duidelijk mogelijk te beschrijven.

## Conclusie

In dit proefschrift heb ik de rol van *common ground* in taalproductie onderzocht door middel van eye-tracking en fMRI. Ik vond onder andere dat *common ground* een vroege invloed heeft terwijl je plant wat je wilt zeggen en dat extra hersengebieden binnen het *mentalizing network* actief zijn wanneer je rekening houdt met *common ground*. Deze resultaten maken duidelijk dat een complete theorie over de neurobiologie van taal ook rekening moet houden met communicatieve en sociale factoren.



## **Curriculum Vitae**

Flora Vanlangendonck was born on August 17<sup>th</sup> 1987 in Jette, Belgium. She studied Linguistics & Literature at the Vrije Universiteit Brussel, and graduated with greatest distinction in 2009. She then enrolled in the Research Master Cognitive Neuroscience at Radboud University in Nijmegen, specializing in psycholinguistics. For her Master's thesis, she investigated bilingual visual word recognition under the supervision of Prof. Ton Dijkstra and Dr. Shirley-Ann Rueschemeyer. She graduated cum laude from the program in 2011. She then started the PhD project that resulted in this thesis at the Donders Centre for Cognitive Neuroimaging, supported by a Donders TopTalent PhD grant. In this project, she studied the neural mechanisms of communicative language production using fMRI and eye-tracking. She was supervised by Dr. Laura Menenti, Dr. Roel Willems and Prof. Peter Hagoort. In 2014, she worked as a policy officer at the Netherlands Organisation for Scientific Research (NWO) for six months as part of the Professional PhD Program organized by the PhD candidates Network of the Netherlands. She finished her PhD thesis in 2016 and currently works as a grant officer at the Research Support Office of the Faculty of Humanities at Utrecht University, where she combines her interests in science and science policy.

## Publication list

Vanlangendonck, F., Willems, R. M., Menenti, L., & Hagoort, P. (2016). An early influence of common ground during speech planning. *Language, Cognition and Neuroscience*, 31(6), 741-750.

Peeters, D., Vanlangendonck, F., & Willems, R. M. (2012). Bestaat er een talenknobbel? Over taal in ons brein. In *Alles wat je altijd al had willen weten over taal: De taalcanon* (pp. 41-43). Meulenhoff.

## In preparation

Vanlangendonck, F., Willems, R. M., & Hagoort, P. (submitted). Taking common ground into account: Specifying the role of the mentalizing network in communicative language production.

Vanlangendonck, F., Takashima, A., Willems, R. M., & Hagoort, P. (submitted). Collaborative learning engages different memory retrieval networks.

Vanlangendonck, F., Willems, R. M., & Hagoort, P. (in preparation). The role of the mentalizing network in the establishment of conceptual pacts.





## Acknowledgements

This PhD has been an amazing journey and there are many people who have helped me along the way. I am very grateful to everyone who has accompanied me on this crazy ride, but more on that below...

Let me start with the people who contributed directly to the scientific work. **Peter**, thank you for creating a research group in which young researchers have the freedom to pursue their own research interests. I admire your sharp mind, your broad scientific knowledge and your incredible dedication to both the DCCN and the MPI. I greatly valued your advice and feedback on my research plans and manuscripts. **Roel**, you became my daily supervisor a year after the start of this project, but the transition went very smoothly. I learned a great deal from you about experimental design and effective writing. Thank you especially for reminding me of the bigger picture when I got stuck on the details and for your confidence when I took a 6-month break to work at NWO. **Laura**, thank you for your enthusiasm and your excellent scientific advice, which helped enormously to get this project on the rails during the first year. I really enjoyed our collaboration and I was sad to see you leave. **Atsuko**, I would not have dared to dive into the unknown field of memory research without you. Thank you for sharing your fMRI and memory expertise with me. I really enjoyed our collaboration. **Dan**, thank you for your patience and advice when I struggled with the fMRI analysis of the third chapter. I hope you are enjoying life in Chicago to the fullest! **Miriam**, thank you for your R advice, e-mail updates about interesting papers and brainstorm sessions about experimental design. You are a warm person and a fantastic scientist. Finally, I am very grateful to the over 250 **participants** who participated in my experiments. This thesis would not have been possible without you!

I am grateful to the **Neurobiology of Language Group** for their excellent advice and feedback on the work in this thesis and for the fun social events. I especially want to thank my fellow PhD students Alina, Ashley, Bohan, Dan, Danchao, Evelien, Franziska, Gwilym, Huadong, Irina, Izabela, Jana, Johanne, Jolien x 2, Lotte, Marvin, Matthias, Nietzsche, René, Richard, Salomi, Sarah and Valeria for the interesting discussions and emotional support during our PhD meetings, dinners and lunches. I also owe a big thank you to the postdocs in our group: Alex, Alexis, Anika, Anne, Atsuko, Chu, Dan, David, Diana, Geertje, Hartmut, Jan-Mathijs, Julia, Karl-Magnus, Katrien, Kirsten, Lennart, Lin, Monique, Roel, Tineke, Vicky, Yingying and Zheng. You made our lab meetings a nurturing and positive place for young researchers to present their ideas.

I would like to thank the **Donders Graduate School** for giving me a TopTalent grant, which allowed me to develop and work on my own PhD project. This was definitely not always the easy road, but it made the journey all the more interesting. Rachel, Dirkje, Kevin and Els, thank you for the support from the **International Max Planck Research School for Language Sciences** over the years.

This thesis would not have been possible without the amazing support of the **technical and administrative groups** at the DCCN. Special thanks to Marek, Mike and the mini-Mareks for replacing my hard disks several times (sorry about that!). Jessica, Uriel and Erik, I am very grateful for your help setting up my crazy multi-screen setups in the behavioral labs. Paul, king of Paul's Pleasure Palace, thank you for teaching me how to operate an MRI scanner, for fixing the scanner during the weekend and for endless entertainment during long scan sessions. Tildie, you helped me so much by organizing meetings with Peter, chasing the manuscript committee and generally making sure the DCCN does not fall apart. Arthur, thank you for making sure we got to enjoy an epic Dagje Uit every year and for carefully selecting the Donders Zevenheuvelenloop teams. I am also very grateful to the friendly ladies at the reception, Nicole, Sandra and Ayse, for making the administration run so smoothly and for dealing with the hundreds of participant forms I submitted over the years! Finally, thank you Lucia for adding extra cheerfulness to the DCCN!

Nietzsche and Dan, thank you for agreeing to be my **paranymphs**. Nietzsche, I am very glad we shared an office in the last stressful months of our PhDs, during which we looked for jobs, complained a lot (sorry Xu!) and worked our asses off. I am sure you will do great in your new job and wish you all the best in Amsterdam! Dan, master of the layers, I am honored that you are honored to be my paranymph ☺ I was very happy that you joined our small but strong NBL delegation at the DCCN. Thank you for organizing our weekly PhD lunches and for fun conversations at the FAD. I wish you the best of luck with the rest of your PhD!

The many Donderians made my time at the DCCN very enjoyable. Less than a month after I started my PhD, I had the honor of joining the awesome **Donders band**. Miriam, Tom, Hugo and Stephen, thank you for the great band practices and concerts. It is high time for a reunion tour! Also a big thanks to my fellow members of **Rank Deficient**, Sean, Ruben, Tom, René, Richard and Ruud, for the weekly entertainment on Wednesday nights and occasional pub quiz victories. My fellow **karaoke** organizers, Verena and Jeanette, I am glad we got this tradition started and hope many more editions will follow. I also greatly enjoyed the many Donders running events, random cake days, random suit days, Berendonck barbecues, Dagjes Uit and MadNes trips. The highlight of the week was always the **FAD**, expertly organized by the one-and-only Donders poet Sean. The list of lovely people I should thank for these great starts to the weekend includes, but is not limited to Alex, Alina, Anke Marit, Anne, Annelies,

Ashley, Corina, American Dan, English Dan, Eelco, blonde Eelke, brunet Eelke, Elexa, Frank, Frauke, Hanneke, Ian, Isabella, Jeanette, Jeff, Jen, Jeroen, Jim, Joost, Lennart, Lieneke, Linda, Lisa, Loek, Lorijn, Marieke, Marlieke, Matthias, Merel, Miriam, Mirjam, Nietzsche, Payam, Peter, René, Ricarda, Richard, Roemer, Ruud, Sander, Sasha, Sean, Stephen, Susanne, Sybrine, Thomas, Tim, Verena, Vincent and Winke.

As a result of the Donders renovation, I have had the pleasure of sharing an office with Loek, Linda, Sander, Tom (except when you threw stuff at me 😊), Lieneke, Reinout, Anke, Nietzsche, Xu, Payam and Zahra. Thank you for being such great **office mates!**

Ik wil graag mijn **nieuwe collega's** in Utrecht bedanken. Brigitte, Hetty, Judith, Martijn en Quirijn, bedankt voor het warme welkom bij het RSO! Ik voel me al helemaal thuis onder de Domtoren 😊

Ik wil ook graag Paul Berendsen en mijn **MaGW**-collega's bedanken voor de mogelijkheid om een half jaar van het werken buiten de wetenschap te proeven. Dankzij mijn NWO-tijd kon ik met hernieuwde motivatie en energie dit proefschrift afmaken. Ik hoop dat jullie het PNN-project zeker verder zetten!

Gelukkig kon ik ook buiten het Donders rekenen op geweldige **vrienden**. Annemieke, Dieter, Lieneke, Marjolein, Hein, Anne, Jenny en John, bedankt om deze PhD-periode van veel wijn, kikkers, kaasplanken, negerzoenen en ander leuks/lekkers te voorzien. Jullie waren een fantastisch support team op weg naar de finishlijn! Nele, Antoine, Sarah en Frederik, bedankt voor de gezellige bezoeken in Brussel, Gent en Nijmegen. Tom, we begonnen samen aan dit avontuur in Nijmegen en ik hoop dat we ondanks de afstand contact blijven houden.

Jessica, Fred, Amanda, Birgit, Julia, oma Atsje en oma Rietje, bedankt voor het warme welkom in jullie **familie**. Ik ben blij dat ik nog mocht terugkomen nadat jullie bij mijn eerste bezoek al meteen vijf vriendinnen op de koffie kregen. Het is erg fijn om een Nederlands *home away from home* te hebben in Flevoland.

Ik wil ook graag mijn tantes, nonkels en de hele **familie** Raeymaekers bedanken voor het organiseren van de jaarlijkse zomerkerst, zodat ik mijn Belgische roots niet vergeet. Een familiefeest is uiteraard niet compleet zonder een varken/geit aan het spit of een gigantische pan paella!

Lieve **mama en papa**, jullie hebben me altijd door dik en dun gesteund. Bedankt voor de wekelijkse Skypegesprekken, jullie vele bezoeken aan Nijmegen en de leuke kerstvakanties. Papa, met een vader die zijn vrije tijd besteedt aan onderzoek naar mossen is het misschien niet verwonderlijk dat ik wetenschapper ben geworden. Het deed veel deugd om te kunnen praten met een PhD-ervaringsdeskundige en ik heb erg genoten van onze mooie

fietstochten door Nederland. Mama, bedankt dat je me hebt laten opgroeien in een huis vol boeken, kunst en cultuur. Van jou heb ik geleerd om buiten de gebaande paden te denken, altijd met een kritische blik naar de wereld te kijken en zo veel mogelijk op eigen benen te staan.

Lieve **Ruben**, waar moet ik beginnen? Ik ben zo blij dat we elkaar hebben ontmoet en ik kan je niet genoeg bedanken voor je steun en geduld. Je stond altijd voor me klaar met een luisterend oor, een peptalk of een portie advies. Samen vierden we onze eerste publicaties en nieuwe banen, en samen worden we dit jaar doctor. Ik ben benieuwd waar ons pad heen zal leiden, maar heb er alle vertrouwen in dat het een mooie reis wordt met jou.

## **Donders Graduate School for Cognitive Neuroscience**

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:

<http://www.ru.nl/donders/graduate-school/phd/>

