# The Contribution of Memory to Common Ground Effects during Language Comprehension

by

**LIN ZHAO**

A thesis submitted to the

University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

May 2017

# Abstract

This thesis presented ten experiments investigating the role of working memory and long-term memory in forming, storing and using representations of what is known (i.e., common ground) among people engaged in communication. Chapter 1 provided a general review of common ground and perspective-taking effects in referential communication. Chapters 2 and 3 examined how memory loads and memory capacities constrain adults and children's ability to use a speaker's perspective in language comprehension. Experiment 1 employed eye-tracking with adult participants, and indicated dissociable roles of working memory and long-term in perspective encoding and perspective integration. Experiments 2-3 observed an age-related improvement in the use of perspective information in language comprehension between 8- and 10-year-olds. Chapters 5 and 6 explored whether effects of common ground could be achieved via a low-level memory-based mechanism reviewed in Chapter 4, without necessarily going though explicit inferences about perspectives. Experiments 4-7 tested whether partner-specific effects could be achieved via memory associations between conversational partners and referents. Experiments 8-10 explored whether an object being in common ground or privileged ground during a preceding discourse would influence people's memory for this object. Finally, Chapter 7 provided a brief general discussion of the findings, and suggested some potential future directions.

# Acknowledgement

Firstly, I would like to thank my two amazing supervisors, Prof Ian Apperly and Dr Jessica Wang, for their professional academic guidance, for their critical and constructive feedback on my work, and most importantly, for their heart-warming encouragement and support. Thank you both for always being patient with me, and for empowering me to achieve my academic potentials. Ian, thank you for being keen to taking me as your PhD student four years ago, when I was still a student majoring in Hindi language and literature. Jessica, thank you for mentoring me in school testing and eye-tracking, and for sharing with me your valuable career advice.

I would like to thank all the participants in my research project for their time and effort. I'm also grateful for the teachers who accommodated my research in the following six primary schools: Cofton Primary School, Forestdale Primary School, Honeybourne First School Academy, Pebworth First School, St. John Fisher Catholic Primary and Water Mill Primary School.

Thanks to my friends who have made my past four years at Birmingham memorable. Thank you for your company and encouragement, and for the cups of coffee, trips, and hot pots we have had together. Thanks to my colleagues in Frankland 513 and 309, as well as other members of Ian's lab, for the stimulating discussion and ideas. Thanks to Olwyn Lai and Na Li for their patience in reading some chapters of the thesis. A special thank you to David Zhou for formatting my thesis, and more importantly, for tirelessly instilling confidence in me when I doubt myself.

Lastly, thanks to my family, especially my parents, for being supportive in every aspect of my life. I could not have embarked on this incredible journey without their support.

## List of Figures

## List of Tables

# CHAPTER I

# General introduction

## 1.1. Overview

Human beings are social. We navigate our social lives largely through communication. At home, we communicate with family members to show how much we love and care about them. At a party, we communicate with friends or strangers to show friendliness and interest. In a seminar, we communicate with other attendees to share our knowledge and understanding of the topic. In business, we communicate with clients to tailor our services and products for their needs. In politics, we communicate with voters to win over their vote. These examples can go on for pages, as we communicate with different people, in different ways, under different situations, driven by different motivations to achieve different goals in everyday life.

Communication is an action jointly achieved by both the speakers and the listeners. Grice (1957), a well-known philosopher of language, has pointed out that successful communication does not merely involve encoding and decoding of information, but also dependent on the ability of making pragmatic inferences based on communication partners' intentions and knowledge. Grice (1975) has proposed the *Cooperative Principle*, which suggests that partners in conversation should act cooperatively in order to convey and understand the meaning of their utterances, and to achieve particular communicative goals. According to this principle, conversational participants 'expect each other to make a conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange' (Grice, 1975, p. 45).

Following the line of Grice's arguments, Clark and Marshall (1978, 1981) have proposed the concept of '*common ground'*, which refers to the mutually shared information between the speakers and listeners. Clark and his colleagues have argued that speakers and listeners should work within their common ground in order to achieve successful communication (e.g., Clark & Marshall, 1978, 1981; Clark & Wilkes-Gibbs, 1986).The concept of ‘common ground’ is not uncommon in our everyday communication. As experienced communicators, we need to be aware of whom we are speaking to and what we share in common with them, in order to tailor our speech and interpretation accordingly. For example, when we are talking to colleagues, we tend to use technical terms to show professionalism; when we are talking to a layman, however, we try to avoid jargons. Much work has been conducted in recent decades by psycholinguistics, social psychologists, cognitive psychologists and developmental psychologists, in an attempt to address some essential questions about common ground. For example, researchers have been exploring *when and how* conversational partners make appropriate inference about common ground between themselves, and *when and how* relevant common ground information is used to guide their communication.

One important line of research in this area, which is the focus of this thesis, examines the role of memory in the inference and use of common ground during communication. Memory has a key role in many aspects of the processes by which common ground guides our communication. For example, during communication we encode in memory our conversational partner’s names, voices, physical features, occupations, what they tell us, and what we have in common with them, whether intentionally or unintentionally. We also retrieve the previously encoded information of such kind from our memory, and integrate it with the messages that we are delivering or receiving. The aim of this thesis is to examine the memory processes involved in inferring, retaining and using common ground information during communication.

This chapter offers a brief overview outlining the historical and current debates surrounding when and how people infer and use common ground information to guide communication. Much of the debates is informed by the literature on theory of mind and perspective taking. This chapter also provides a rationale for examining the relationship between people's ability to infer and use common ground information and the underlying memory processes within the context of referential communication. Importantly, the overview here is not meant to be an exhaustive or comprehensive literature review of the research on common ground in communication. Instead, it is mainly to introduce readers with the research context in which this thesis is based. Please note that 1) Chapter 2 and Chapter 3 each contains an introduction section which reviews literature relevant to that chapter; 2) Chapter 4 has been written as an introductory chapter to Chapters 5 and 6. In Chapter 4, relevant research on the low-level memory-based mechanism of common ground has been reviewed, and the rationale for the research questions of Chapters 5 and 6 has been set up.

## 1.2. Common ground and perspective taking in referential communication

According to the classic view proposed by Clark and Marshall (1978, 1981), common ground can be established on three bases: *physical co-presence*, *linguistic co-presence* and *community membership*. In other words, mutual knowledge between conversational partners can be obtained through shared physical environment, shared past / present conversations and shared sociocultural background between speakers and listeners. It is worth noting that these three types of co-presence are heuristics, which communicators draw upon to assess whether one piece of information is in the common ground. This account does not provide an operational definition of common ground, though.

In order to form the assumptions of common ground for communication to work successfully, speakers and listeners need to appreciate the perspective of another person. People bring their own, usually unique, perspectives into communication, such as different beliefs, desires, intentions, knowledge and experience. In our everyday life, people's differing perspectives usually influence the manner and content they communicate with each other.

*Perspective-taking* ability, as an essential ability in human social cognition, is involved substantially in the representations of common ground. Previous research has examined perspective taking along two dimensions. One is visual perspective taking. This concerns how people take into account another individual's visual point of view when it differs from one's own, and how this perspective difference influences what people see and talk about (e.g., Apperly et al., 2010; Keysar, Barr, Balin, & Brauner, 2000). The other dimension is mental perspective taking, which concerns how people reaches the understanding that other individuals' behaviours are guided by their own distinct mental states, such as their beliefs, desires or intentions (e.g., Ferguson, Apperly, Ahmad, Bindemann, & Cane, 2015).

A large number of studies examining when and how people attend and use conversational common ground in communication has been conducted within the context of *perspective taking* in *referential communication*. These studies examine how speakers took their listeners' perspectives into account and adapted their way of referring to a specific item or a piece of information accordingly (e.g., Brennan & Clark, 1996; Horton, 2007; Nadig & Sedivy, 2002; Nilsen & Graham, 2009; Wardlow, 2013). These studies also examined how listeners took their speakers' perspectives into account and adjusted their interpretation of the speakers' referring expressions accordingly (e.g., Apperly et al., 2010; Brown-Schmidt, 2009a, 2009b; Heller, Grodner, & Tanenhaus, 2008; Keysar et al., 2000; Keysar, Lin, & Barr, 2003; Lin, Keysar, & Epley, 2010; Nadig & Sedivy, 2002; Nilsen & Graham, 2009).

The representations of common ground play an essential role in shaping the production and comprehension of referring expressions. For example, one interesting phenomenon is that when one person has previously established a referring term for a specific item with a conversational partner, both people tend to use the same term repeatedly to refer to the same object. They also tend to expect each other to use the same term. This phenomenon is termed as 'lexical entrainment' (e.g., Brennan & Clark, 1996; Garrod & Anderson, 1987), which will be further explained in details in section 4.3 in the thesis. This phenomenon exemplifies how people use mutual knowledge and experience to tailor their way of referring to items. To give another example, if people refer to an item whose name (a novel meaningless label) has been previously learnt together with their conversational partner, they tend to use the name itself to refer to the item. In contrast, if people refer to an item whose name is not shared with their conversational partner, they tend to include detailed descriptions to help their conversational partner identify the item (e.g., Heller, Gorman, & Tanenhaus, 2012). Another example is that when people hear an instruction 'pick up the glass' from a speaker, they tend to choose the glass which is visible to the speaker even if there are two glasses from their own point of view. However, when both glasses are visible to themselves as well as the speaker, people tend to ask for further clarification (e.g., Nadig & Sedivy, 2002). These examples demonstrate that people use common ground information to shape how they produce and interpret referring expressions in communication.

There is little doubt that common ground information modifies the language that we produce and shapes how we comprehend the speech of others. What remains unclear, however, is when and how people make use of common ground information during communication. The following two sections in this chapter will review relevant literature on when and how people access common ground in communication respectively. Please note that the accounts regarding the time course of common ground effects in communication also provide explanations of the underlying mechanism of how common ground affects language use. The reason why I group the theories into two separate sections regarding when and how people access common ground is to highlight the two specific focuses of the intensive debates surrounding the use of common ground in the existing literature.

## 1.3. Different accounts of when people access common ground in communication

There has been an intensive debate on the time course by which common ground information is incorporated in language use during communication. The invention and increasing use of eye-tracking devices allow researchers to monitor participants' eye movements during moment-by-moment language processing. Much research has been carried out by adapting the *visual world paradigm* (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The visual world paradigm has been mostly used to study people's online language comprehension. In a typical visual world experiment on language comprehension, participants are presented with dynamic visual scenes as the verbal messages unfold, and participants' moment-by-moment eye movements are recorded and analysed to reveal the online time course of their language processing. For example, Altmann and Kamide (1999) presented participants with a visual display showing a boy, a cake and some other non-edible objects. Participants was presented with either the sentence 'they boy will move the cake' or 'the boy will eat the cake'. They found that participants were quicker to launch a fixation on the cake in the display when the word 'eat' was spoken compared to when the word 'move' was spoken. Furthermore, in the 'eat' word condition, participants fixed on the cake before the word 'cake' was mentioned. However, in the 'move' word condition, participants fixed on the cake after the word 'cake' was heard. This pioneering study demonstrated that participants were able to integrate the linguistic information with the visual display to identify the domain of reference.

The eye-tracking visual world paradigm provides rich continuous data without intruding into participants' performance. Several types of measures are available for analysis. For example, researchers can compute the total dwell time or the number of fixations on certain objects or locations in the visual display during a fixed window of time. They can compare the likelihood of participants' look at the target object and that of competitor objects, and observe how this likelihood changes over time. They can also calculate the amount of time it takes for participants to launch their first fixations or final fixations on certain objects or locations. With this visual world paradigm, researchers have provided some empirical evidence for accounts regarding the time course of the effect of common ground in language use. Four different accounts are outlined in the following sections.

### 1.3.1. Reference diary account

Clark and Marshall (1978) propose that people use *Reference Diaries*, i.e., an updated record of partner-specific knowledge, to help them keep track of common ground. People consult their partner-related reference diaries to determine whether a piece of information is shared or privileged, and whether the information can be used for guiding communicational references. According to this theory, the representations of common ground are rich, diary-like and episodic. People search in their memory for an episodic event which evidences physical, linguistic and/or cultural co-presence. More importantly, the reference diary account suggests that '*we must search in every case for an event (of co-presence)*' (Clark & Marshall, 1978, p. 63). This implies that common ground acts fully as the domain of interpretation, and that it serves as an immediate constraint upon language production and comprehension. The reference diary account provides an analogy between how people assume common ground and how people consult a reference diary. However, it is unclear how exactly this reference diary functions.

### 1.3.2. Perspective adjustment model

A series of studies conducted by Keysar and his colleagues (e.g., Keysar et al., 2000; Keysar, Barr, Balin, & Paek, 1998; Keysar et al., 2003) provide compelling evidence against the reference diary theory. These studies employed a communication task (i.e. the director task) to examine perspective-taking performance in the context of referential communication. In a typical director task, participants were asked to follow a director's instructions to move the referred objects around a grid. Some objects were visible to participants and the director, thus in the common ground, while some other objects were occluded from the director's perspective, thus in participants' privileged ground. Participants had to rule out the distractors in the privileged ground, which competed against the targets as potential referents, and move the target referents in the common ground instead. However, findings revealed that participants occasionally processed director's instruction using their privileged knowledge, incorrectly moving the item that was invisible to the director. Eye-tracking results further indicated that participants frequently looked at the distractor in their privileged ground before they moved the target object in the common ground. Such egocentric bias was found at the earliest moments of language processing.

Based on these findings, Keysar and colleagues propose the *perspective adjustment model* (Barr & Keysar, 2002; Keysar et al., 1998). According to this model, two processes operate sequentially during language comprehension. At the initial stage, people interpret another person's utterances without taking common ground into consideration. This initial process is relatively efficient and effortless, and it is followed by a relatively slow and effortful adjustment process that is sensitive to common ground information. Based on the perspective adjustment account, common ground only act as a second-stage filter which detects and corrects any egocentric and faulty interpretations established in the initial stage, rather than acting as an immediate constraint on language processing. Consequentially, if people fail to use common ground information to adjust the initial egocentric interpretation, communication will suffer from egocentric errors. Epley, Keysar, Van Boven and Gilovich's (2004) study also provided supporting evidence for the perspective adjustment model. This study found that it took participants a longer time to recognise a different perspective from another individual than a similar perspective. It was also observed that participants' degrees of egocentrism increased under time pressure, and decreased with motivation or rewards for accuracy. This implies that it took efforts for participants to adjust an initially established egocentric interpretation. According to Epley et al. (2004), these findings suggest that people initially anchor their interpretation of utterances from a speaker on their own perspective, before using information about the speaker's perspective to account for differences between themselves and others and to incrementally adjust away from the anchor.

### 1.3.3. Constraint-based model

Some studies, however, suggest that people are not always initially egocentric. For instance, Hanna, Tanenhaus, and Trueswell (2003) found that although listeners failed to completely ignore the interference of privilege ground, common ground still had an immediate effect on reference resolution. The first Experiment in Hanna et al. (2003) employed a referential communication task similar to the director task in Keysar et al.'s study (2000, 2003). Participants were asked to follow a director's instructions (e.g. 'Put the blue triangle on the red one'). The target object was the intended referent which was visible to the director (e.g., the red triangle in the common ground), while the competitor was the object which shared the same features with the target but was invisible to the director (e.g., the red triangle in the privileged ground). Although participants could not fully ignore the competitor in the privileged ground which also matched the director's description, they still showed preferential looking at targets over competitors from the earliest stage of reference resolution.

Based on the results above, Hanna et al. (2003) have proposed that common ground could be regarded as a probabilistic constraint within the framework of a *constraint-based model* (e.g., MacDonald, 1994; Tanenhaus & Trueswell, 1995). Therefore, different types of information, including not only common ground information, but also other factors (e.g., phonological information, semantic information and previous discourse), can all act as probabilistic constraints and can be tracked in parallel. People integrate information from different sources to generate a most plausible interpretation. Whether a certain piece of information succeeds in the competition with others is dependent on its salience and strength in a specific context. This account implies that the effect of common ground information varies with its salience and strength. For example, if the likelihood for an item to be mentioned by a conversational partner seems quite irrelevant to whether the item belongs to common ground, then the common ground information is not a very salient and consistent constraint. In this case, listeners might not exhibit strong effects of common ground when they communicate with the speaker. The constraint-based model, therefore, is partially consistent with the perspective adjustment model, in that common ground does not completely constrain the domain of interpretation for referring expressions. However, it argues against the perspective adjustment model's idea that common ground is fully ignored in the earliest moments of language processing and that it can only affect later adjustment.

A growing body of research supports the constraint-based model by providing evidence that common ground partially modulates the initial stage of language processing (e.g., Brennan & Hanna, 2009; Brown-Schmidt, 2009a, 2012; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Heller et al., 2008). For example, Heller et al. (2008) used a variation of the director task with a group of four items the display. Two factors were manipulated in their study. First, the group contained either one pair of objects with contrasting sizes (e.g., small duck, big duck, box, and soap) or two pairs (e.g., small duck, big duck, small box and big box). Secondly, the soap in the 'one contrast' condition and the small box in the 'two contrast' condition were either visible or occluded from the speaker's perspective, while all other three objects were visible to the speaker. Participants heard instructions such as 'pick up the big duck'. This study found that participants were quicker to disambiguate the referent in the 'one contrast' condition than the 'two contrasts' condition, which demonstrated that participants expected the referent to be from a pair of items with contrasting sizes when hearing the scalar adjective (e.g., 'big'). More importantly, in the 'two contrasts' condition, participants showed an earlier disambiguation of referent when the second pair of contrasts contained an object which was occluded from the speaker's perspective, in contrast to when all two pairs were visually shared. Heller et al. (2008) concluded that participants were able to use perspective information from the early moments of language processing.

### 1.3.4. Anticipation-integration account

Barr (2008) employed a variant of the director task, where instructions contained temporary referential ambiguity. Among the four objects displayed on the grid, two objects were phonological competitors (e.g., a bucket and a buckle) which made the target of the speaker's referring expression temporarily ambiguous until they heard the last part of the object name (e.g., 'et' or 'le'). Compared to when neither of the two objects were occluded from the speaker's perspective, participants were more likely to fixate on the object in the common ground upon hearing the first part of the instruction (e.g., 'click on the') when one of these two objects was occluded from the speaker's point of view. However, participants were equally likely to fixate on the object in the privileged ground and the object in the common ground when hearing the first part of the object name 'buck'. Such dissociation in the effects of common ground in different phases demonstrated that although participants did anticipate speakers to refer to an object from common ground, they were unable to reduce the egocentric interference in a later phase when they needed to integrate linguistic input with the perspective information.

Based on these findings, Barr (2008) has proposed the *anticipation-integration account*. According to this account, people may have an early anticipation for common ground information prior to the point when they need to integrate perspective information with linguistic inputs. Nonetheless, the later integration process still suffers from the autonomous activation of egocentric knowledge, despite participants' explicit awareness of the need to take another person's perspective into account. Therefore, the mechanism underlying perspective taking not only involves the process of using common ground information, but also involves ignoring the autonomously activated egocentric interference during the integration phase. Unlike from the perspective adjustment model, this account suggests that listeners do not strategically or adaptively 'anchor' on their own egocentric perspective. Instead, the anchoring effect is imposed upon them by a relatively low-level autonomous activation of privileged knowledge, which reflects human's processing limitations.

## 1.4. Different mechanisms of how people access common ground in communication

### 1.4.1. Literature on theory of mind

The accounts of common ground often assume that people rely on the theory of mind ability to infer another person's perspective. Theory of mind (which is also termed as mentalising, mindreading, or folk psychology) is the ability to understand and reason about other people's mental states such as their beliefs, desires and intentions (Astington & Gopnik, 1991; Perner, Leekam, & Wimmer, 1987; Premack & Woodruff, 1978; Wimmer & Perner, 1983). Theory of mind is believed to play an essential role in the cognitive processes of communication. For example, Grice (1957, 1975) has argued that human communication features the importance of making pragmatic inferences about each other's intentions. Building upon Grice's view, Sperber and Wilson (1986, 2002) have proposed that a specialised 'theory of mind module' or a 'dedicated inferential mechanism' is involved in the interpretation processes of communication. This theory of mind module provides mindreading inferences, which help listeners to generate the most relevant interpretation and also help speakers to produce the most appropriate utterances.

Early and traditional research on theory of mind has focused on the question of when young children start to gain such theory of mind abilities by passing the classic false belief tasks (e.g., Astington & Gopnik, 1991; Perner et al., 1987; Wimmer & Perner, 1983). A meta-analysis study suggests that children develop the necessary 'theory of mind' concepts to be able to successfully infer other people's mental states between the age of 2 to 7 years old, and mostly begin to pass the classic false belief tasks between 3 to 5 years old (Wellman, Cross, & Watson, 2001). Recent years have seen an increase in the number of studies on older children and adults who have grasped the basic 'theory of mind' concepts, as well as younger infants whose cognitive abilities and social experience are very limited. Surprisingly, some studies have found that older children and adults, who have presumably obtained the basic understanding of mental states, still experience difficulty in using theory of mind ability to accommodate another's perspective which is different from their own (e.g., Apperly et al., 2010; Dumontheil, Apperly, & Blakemore, 2010; Keysar et al., 2000, 2003). On the other hand, some studies have also showed that when being examined by indirect measures such as looking time instead of being asked to provide explicit verbal responses, even infants seem to be sensitive to other people's mental states (e.g., Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007).

To reconcile these seemingly conflicting evidence, Apperly and Butterfill (2009) have proposed that there are two distinct systems of theory of mind. One system involves explicit inference of others' mental states. This system allows the theory of mind reasoning to be flexible and accurate, but this flexible reasoning process is limited by its heavy demands on general processing resources, such as language, memory capacity and inhibitory control. This idea of a controlled but effortful theory of mind system is consistent with the finding that even adults might experience difficulty in using theory of mind to accommodate another person's differing perspective, especially when their general processing resources were occupied by other cognitively demanding tasks (e.g., Lin et al., 2010). It is also consistent with the finding that children between 3-5 year olds started to pass the verbal false-belief tasks as their language and executive functions develop (Carlson, Moses, & Breton, 2002; Carlson, Moses, & Claxton, 2004; Wellman et al., 2001). The other system, in contrast, allows the theory of mind reasoning to be relatively automatic, efficient and effortless, so that one can make quick inference about another's mental states in order to go through moment-by-moment social interaction smoothly. However, this system is inflexible and may not be accurate on complex social occasions. According to Apperly and Butterfill (2009), the non-verbal studies on infants, which showed that infant participants' eye movements were sensitive to another's beliefs (e.g., Onishi & Baillargeon, 2005), provided supporting evidence for such an implicit theory of mind system.

### 1.4.2. Dual-mechanism account of common ground

The idea of a dual-system model is not uncommon in the research field of social cognition (see Chaiken & Trope, 1999, for a review on the dual-process models of social information processing). Evidence from some cognitive areas such as number cognition (Feigenson, Dehaene, & Spelke, 2004) also support the notion of a dual-process model. The dual-system models proposed by different researchers typically share one feature: one system is a relatively high-level domain-specific system which operates under deliberate control and relies on executive resources, whereas the other is a relatively low-level domain-general system which operates without conscious control and is relatively automatic, efficient and effortless (e.g., Kahneman, 2011).

Similarly, such a dual-mechanism system may both exist in the process of using common ground information during communication. There may be a high-level mechanism, which allows people access common ground information by inferring other people's mental states in a controlled and effortful way. There may also be a low-level mechanism, which enables the use of common ground in communication to be achieved through domain-general processes, in an automatic and effortless way, without any necessity to go through explicit mentalising. In the following two sections, I will explain how each of these two mechanisms may operate to help people access common ground information during communication. I will also discuss the contribution of memory to each mechanism.

### 1.4.3. High-level mechanism: explicit mentalising

In some circumstances, a high-level controlled and effortful mechanism may underlie the processes of using common ground during communication. Specifically, people may get access to common ground information by making deliberate inference about another conversational partner's perspective, and then use the relevant common ground information to guide communication. People do explicitly judge another conversational partner's perspective in some social situations. For example, when your girlfriend is clearly angry at you, but only says 'nothing' when you ask her what has happened; or when your supervisor hesitates for a while and then says 'interesting' after you propose a research idea to him at the supervision meeting. If you encounter these social situations for the first time, you will need to explicitly draw evidence from different sources, to make efforts to figure out what the other person is actually thinking about, and to interpret their messages ('nothing' or 'interesting') accordingly. The previously-mentioned 'director task' (e.g., Apperly et al., 2010; Keysar et al., 2000, 2003) also introduces such a complex social situation in which a director's perspective differs from participants' own. Participants are explicitly required to make inference about the director's limited perspective and then use inferred perspective information to identify the target referent. If participants do not accommodate the director's perspective, they will exhibit misinterpretation of their partner's messages.

As mentioned earlier, Apperly and Butterfill (2009) has argued that individuals' explicit mindreading abilities are limited by their executive functions resources, one key component of which is working memory. Several studies have also been done to examine how working memory capacity and working memory load affect people's ability to infer and use common ground in communication, but the results seem rather mixed. Some studies found that individuals' working memory capacities affected their perspective-taking performance in communication (e.g., Lin et al., 2010; Wardlow, 2013); whereas other studies failed to observe any effect of working memory capacities on perspective taking (Cane, Ferguson, & Apperly, 2016; Nilsen & Graham, 2009). The mixed results of these studies will be explained in details in the introduction section of Chapters 2 and 3. Experiments in Chapters 2 and 3 employed an adapted version of the director task to unpack the following three questions. Firstly, apart from working memory, how might long-term memory contribute to inferring and using common ground information in communication? Secondly, what memory factors embedded in everyday communication might influence people's ability to take another person's perspective and to use this perspective information in referential communication? Thirdly, how might different memory processes contribute to children's age-related improvement in perspective taking during communication?

### 1.4.4. Low-level mechanism: submentalising

As explained in the previous section, certain circumstances might require one to explicitly make pragmatic inference about another conversational partner's mental states. However, people cannot fully reply on the deliberate consideration about others' mental states to navigate everyday social situations, as this deliberate mentalising process is usually achieved with a cost of speed and efficiency (Apperly & Butterfill, 2009). Rather, a lot of communicative situations require one to be quick and efficient enough to go through moment-by-moment social interaction with other people. Is it possible for people to successfully access information mutually shared with other people during communication without explicitly reasoning about their mental states?

Heyes (2014) is the first to have proposed the concept of 's*ubmentalising*'. According to Heyes (2014), the notion of *submentalising* is different from the notion of *implicit mentalising*. Implicit mentalising, as a type of mentalising, requires people to interpret another's behaviours by thinking about their mental states in an efficient and effortless way, instead of a slow and effortful way. Submentalising, in contrast, is a domain-general cognitive process which does not involve representing other people's mental states but can simulate the effects of mentalising in social contexts. Heyes (2014) argued that many empirical evidence of efficient and spontaneous perspective taking which was assumed to be due to implicit mentalising, could instead be explained by domain-general cognitive mechanisms via the route of submentalising, such as involuntary attentional orienting and spatial coding. In the thesis, however, I do not intend to take stance in this debate, but rather to use this 'submentalising' concept to illustrate how people may rely on domain-general cognitive mechanisms to produce social behaviours which may look as if they were carried out by taking into account a conversational partner's mental states.

Although the term 'submentalising' only appears in literature recently, the idea that some social effects may be achieved via basic low-level cognitive mechanisms is not unusual. In the literature on common ground, Horton and Gerrig (2005a, 2005b) have proposed an influential account that common ground can be achieved via a *low-level memory-based mechanism*. This memory-based mechanism puts an emphasis on the contribution of domain-general memory encoding and retrieval processes to the use of common ground information during communication. It is proposed that people do not only encode the content of a conversation in their memory, but also encode the contextual information associated with the content, for example, the person with whom they have the conversation. These partner-specific associations are stored in episodic memory traces and can be automatically activated by salient memory cues related to the specific conversational partners (e.g., their physical presence or voices) in working memory, enabling people to retrieve and use common ground representations in an implicit, fast and effortless way. Many instances of partner-specific effect, which may look as if they occurred on the basis of people's deliberate consideration of common ground, can instead be explained by such a relatively spontaneous, efficient and effortless domain-general memory process. More information about this low-level memory-based mechanism of common ground and its supporting evidence have been reviewed with details in Chapter 4, which serves as an introductory chapter to Chapters 5 and 6. Experiments in Chapters 5 and 6 aimed to provide evidence in support of such a low-level memory-based mechanism underlying common ground representations.

In summary, I am proposing a dual-system account about how people access common ground in communication. Under the high-level controlled and effortful mechanism, people make inference about common ground via 'explicit mentalising'. This high-level mechanism depends heavily upon people's cognitive resources, e.g., working memory and inhibition control. In contrast, the low-level mechanism suggests that the use of common ground can often be achieved via 'submentalising'. An important instance of such a submentalising mechanism is the low-level memory-based mechanism of common ground proposed by Horton and Gerrig (2005a, 2005b), advocating that domain-general memory processes can help people achieve the effects of common ground in communication without any necessity to go through explicit mentalising.

## 1.5. The memory processes in discourse comprehension

The main purpose of the present study is to examine the memory processes involved in encoding, retaining and using common ground information in communication. This thesis is particularly interested in how might memory load and memory capacity constrain people's ability to infer and use common ground, and how might domain-general memory processes support the inference and use of common ground information during communication. In this section, I will review some theories on the memory processes involved in discourse comprehension, which might cast light on the questions that this thesis aims to address.

According to a review on discourse comprehension (Graesser, Millis, & Zwaan, 1997), traditional models usually hold the view that there are three memory stores employed in discourse comprehension, namely, short-term memory, working memory and long-term memory. Short-term memory usually focuses the most recent clause being read. Working memory actively holds and processes approximately the recent two sentences in mind. The information actively rehearsed and encoded in working memory will later be stored in long-term memory.

Ericsson and Kintsch (1995) have proposed a concept of ‘long-term working memory’, which is an extended working memory mechanism that allows relevant information in long-term memory to be quickly and reliably retrieved through retrieval cues in the traditional working memory system. As suggested by Ericsson and Kintsch (1995), the mental representations of a situation described by one or two sentences being read are actively held in the focus of attention, and thus in working memory. Once people start a new sentence, the representations of the previously read sentences will be dropped from the focus of attention into long-term memory. Some elements of the actively maintained information in working memory may be associated with certain previously encoded information in long-term memory, and these elements may act as memory retrieval cues to help the relevant information in long-term memory to be efficiently retrieved. This process creates the mechanism of long-term working memory.

## 1.6. Summary

This general introduction to the thesis has outlined several different accounts of when and how people access common ground information in communication, and has reviewed relevant literature which suggests that memory may act as an important factor underlying the representations of common ground during communication. This thesis presents a series of empirical studies, which have employed both behavioural and eye-tracking methodologies, to investigate the memory processes involved in encoding, retaining and using common ground information shared between conversational participants during referential communication. The thesis focuses on two main questions. The first part of the thesis, consisting of Chapters 2 and 3, examines how memory load and memory capacities constrain both adults' and children's abilities to take into account common ground information in language comprehension. The second part of the thesis, including Chapters 4, 5 and 6, explores whether common ground effects could be achieved via the low-level memory-based mechanism proposed by Horton and Gerrig (2005a, 2005b), without necessarily going through explicit theory of mind inference.

CHAPTER II

# Remembering what she knows: The costs of holding in mind a speaker's perspective during a communication task

## 2.1. Introduction

Imagine that you and your Thai friend go to a nice Thai restaurant and sit down at a table. You start to read the menu but find out the menu is completely written in Thai which you have no knowledge about, and there is no picture on the menu to help you order. Suddenly a waiter passes your table, holding a plate of food that looks very appetizing and smells like heaven. Your and your partner's attention is attracted by that plate of food. After the food has been served, you ask your partner, 'I wonder what that is?' You partner does not seem to be surprised by your question, and does not ask what exactly you are talking about, but rather tells you the name of the dish that just passed by. Although in principle your utterance of 'that' is entirely ambiguous, and although the dish that you are referring to is no longer visible to both you and your partner, your partner still successfully interprets your question in the context of the common ground information that has been briefly seen by both of you and thus shared between you.

As reviewed in Chapter 1, much work on common ground has used communication games, in which participants follow or issue instructions to an interlocutor whose perspective differs from their own. For example, in the 'director task' (Keysar et al., 2000) participants were asked to move objects around a grid by a 'director' whose position meant he only saw some of the objects. Objects occluded from the director's point of view remained visible in participants' privileged ground; other objects were visible to both the director and participants, and thus in their common ground. Critical instructions required participants to select a matching referent from the common ground while ignoring a matching object in the privileged ground. Perhaps the most striking observation from such studies is that adult participants, who clearly understand the need to take the director's perspective into account, often fail to do so (Abu-Akel, Wood, Hansen, & Apperly, 2015; Apperly et al., 2010; Keysar et al., 2000, 2003).

It seems plausible that taking account of a speaker's perspective and using this information to rule out distracting alternative objects for their referring expressions may depend upon a participant's capacities for memory and for inhibitory control. However, efforts to test these hypotheses have been met with mixed results. Lin et al. (2010) used a dual task to manipulate the concurrent memory load while participants undertook a version of the director task. These authors also measured participants' working memory capacity using an Operation Span Task (OSPAN task, La Pointe & Engle, 1990; Turner & Engle, 1989), which was believed to measure the ability to maintain multiple items in memory while performing distracting tasks (Unsworth, Heitz, Schrock, & Engle, 2005). Participants under high working memory load, and participants with low working memory capacity, showed longer decision times for selecting a referent that was compatible with the director's perspective. Brown-Schmidt (2009b) found evidence that individual differences in inhibitory control were related to participants' ability to ignore a privileged ground distractor during a director task with temporary referential ambiguity. Wardlow (2013) found that individual differences in both working memory and inhibitory control were related to a speaker's ability to produce utterances adapted to the perspective of their addressee. Nilsen and Graham (2009) found convergent evidence for the role of inhibitory control in children's successful accommodation of a director's perspective, but no role for working memory. Using a variant of the director task, Ryskin, Benjamin, Tullis, and Brown-Schmidt (2015) found no relationship between memory or inhibition and participant's successful use of the director's perspective during language comprehension, but a positive relationship between participants' performance on the working memory task (OSPAN task) and their perspective-taking during production. Cane et al. (2016) used a director task very similar to that used by Lin et al., (2010), but failed to find an effect of individual differences in working memory on listeners' ability to accommodate the director's perspective.

Three limitations of existing studies may help explain these inconsistencies, and more importantly, cast new light on the processes involved in the management of common ground, and how they can be studied experimentally. Firstly, in most existing studies it is not clear how much information participants must actually infer about the director's perspective, or for how long the information must be stored. These tasks typically present participants with a grid in which multiple objects appear in common and privileged ground, followed by a series of instructions referring to items in the grid. One option available to participants is to first identify which items the director can see and remember these as the common ground to which she may refer. However, a second option for participants is to process the director's perspective online during comprehension of her instruction; an option made plausible by evidence that simple visual perspectives can be calculated rapidly and relatively automatically (Qureshi, Apperly, & Samson, 2010; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). In this case, the demands on memory for tracking the director's perspective would be greatly reduced. It is currently unclear which of these options participants pursue, or whether the option pursued varies between individuals or changes across different task variants. In the present study we limited participants' options by removing cues to the director's perspective after participants' initial period of viewing the grid, thereby ensuring that they had to infer and store information about the common ground before the need to use this information in language comprehension.

A second limitation of existing work is the number of objects in the common ground (and thus the potential memory load) has varied across studies. Some studies used either 2 x 2 grids with 1 object per slot (Nadig & Sedivy, 2002), or 3 x 3 grids with 1 object in each of the 4 corners of the grid (Nilsen & Graham, 2009). One out of the four objects was usually blocked off and thus in the privileged ground; the other three objects were in the common ground. Most other studies used 4 x 4 grids where 5 out of 16 slots were usually blocked off as privileged ground and the remaining 11 slots were in the common ground. Seven or eight objects in total were usually presented in these 4 x 4 grids, and some slots in both the common and privileged ground contained no object (Apperly et al., 2010; Cane et al., 2016; Keysar et al., 2000, 2003). The number of items in the common ground has obvious face validity as a variable that will also be relevant in naturalistic communication, and its variation between existing studies may account for some variation in participants' performance and for variation in the role of working memory. The present study will be the first to systematically vary this factor within a single study.

A third limitation is that by focusing exclusively on working memory, as opposed to longer term retention, existing studies may be missing critical demands both in the task and in the management of common ground in everyday communication. Working memory concerns the active manipulation and integration of information over short periods of time and with a storage capacity limit of just a small number of items (Baddeley, 1986, 2003; Baddeley & Hitch, 1974). While it seems plausible that working memory may be important for integrating common ground information with the speaker's current utterance, it is much less clear that working memory would be the main resource for retaining common ground information over an extended discourse. As a point of comparison, text comprehension is generally thought to depend on the use of working memory to integrate information from the small tract of text currently in focus with information from paragraphs or pages earlier, with the latter being stored and retrieved from longer-term memory (Ericsson & Kintsch, 1995; Graesser et al., 1997). The present study overcomes this limitation by examining effects of common ground size both during the period when participants must infer and encode this information and in the period when they must integrate it with the director's instructions. Apart from experimentally manipulating the potential memory load of the task, we also measured participants' working memory capacities using the Operation Span Task (Turner & Engle, 1989), and examined whether effects of common ground size in either time period were related to participants' working memory capacities.

**The present study**

Since our novel design ensured that participants had to process the director's perspective and then hold in memory which objects were in the common ground, we expected to maximise our chances of observing memory effects that have been inconsistently reported in previous studies. Firstly, we expected that when participants had to integrate common ground information with the director's instruction, participants with higher working memory capacity would be less 'egocentric' than those with lower working memory capacity. Please note that being 'egocentric' in the director task means that participants would show faster and more accurate responses on experimental trials where it was necessary to integrate common ground information, compared with control trials where this integration was unnecessary. Secondly, we also expected that when there were more items in the common ground, participants would need to devote more resources to inferring and encoding common ground information when they first viewed the grid, and would be slower and more error prone when they had to integrate common ground information with the director's instruction. Critically, interactions between effects of working memory capacity and common ground size would be informative about whether these factors were related. We were particularly interested in whether the degree to which participants with higher versus lower working memory capacities differ in egocentric tendencies during the integration phase would be greater when common ground was larger.

## 2.2. Method

### 2.2.1. Participants

Thirty-seven native English speakers (15 males, mean age 22.51, range from 19 to 38) from the University of Birmingham participated in our experiment for a small honorarium. One additional participant's data were lost due to technical problems, thus not included in the analysis. Another additional participant was excluded prior to the analysis because of the failure to perform significantly above floor (i.e., failing to respond correctly on more than 3 out of the 16 experimental trials), as it was unclear whether the low accuracy was due to extreme egocentrism, memory failure, or the difficulty in understanding the instructions at the first place.

### 2.2.2. Apparatus

A variant of the director task was presented to participants on a computer screen using the Experiment Builder software (SR Research Ltd, Mississauga, Ontario, Canada). We recorded participants' eye movements with an Eyelink 1000 using the tower configuration (SR Research), recording at 1000Hz. All participants' eye movements were tracked through their left eyes, and their head was kept still with the help of a chin rest and forehead rest. Stimuli were presented on a 24-inch computer screen approximately 60 cm away from participants. Participants' fixations were calibrated via a standard 13-point calibration process before each block. A drift correction (central fixation check on the screen) was also carried out prior to each trial. The grid-images subtended 26.93° (width) by 20.15° (height). We drew interest areas around each slot on the grid, which subtended 3.25° (width) by 3.15° (height). The OPSAN task was run by E-prime 2 software (Psychology Software Tools, Pittsburgh, PA) on an ASUS 14-inch laptop.

### 2.2.3. Design and procedure

**Referential communication task—perspective-taking**

A 2x4 within-participant design was constructed with condition (experimental, control) and the magnitude of common ground (3, 5, 7 or 9 objects) as the two factors.

Each trial consisted of an image of a 4 x 4 grid with 12 objects on it. A female avatar was standing behind the grid. Of the 12 objects on the grid, 3, 5, 7, or 9 objects were visible to both the participants and the female director, and were thus in their common ground. Other objects (correspondingly 9, 7, 5 or 3 in quantity) were occluded from the female director's perspective by green squared backgrounds, and were thus in the participants' privileged ground.

In the instruction phase, participants were informed that the director could not see and could not know about the objects in the blocked slots, therefore would not ask them to move any of those objects. Participants were not only informed of the perspective difference between themselves and the female director, they were also given an example demonstrating how this perspective difference could constrain the director's reference. A previous study found that by giving such comprehensive instructions, adults' proportion egocentric errors were decreased dramatically from 42.2% to 3% (Wang, Cane, Ferguson, Frisson, & Apperly, submitted). An image showing how the grid appeared from the director's point of view was also given to the participants.

A total of 32 grid-images were presented in the test phase. Upon presentation of the grid image, participants had 5000ms to view the image and remember what objects the director does and does not know about. After 5000ms, all 16 slots were occluded by a green background, leaving no visual cues to indicate the perspective difference between the participants and the director. After the occlusion of all slots, the verbal instructions began. Each grid-image consisted of 3-5 instructions including one critical instruction, and the position of the critical instruction varied in each trial with the caveat of never being the last instruction of the trial. A 500ms pause was presented after each instruction. The structure of the filler instructions was either ‘nudge the [scalar/normal adjective] [noun] one slot [directional word]’ or ‘nudge the [noun] one slot [directional word]’; whereas the structure of the critical instructions was always ‘nudge the [scalar adjective] [noun] one slot [directional word]’ (e.g., nudge the small ball one slot up). Directional words included ‘left’ ‘right’ ‘up’ and ‘down’. Participants were told that the directional words ‘left’ and ‘right’ referred to their own left and right sides. For critical instructions, only ‘up’ and ‘down’ directional words were used.

Sixteen grid-images were used for the experimental condition and the control condition respectively. In the experimental condition, the object which best-fitted a critical instruction from the participants’ point of view was always in the participants’ privileged ground, thus making it a ‘distractor’ against the target referent. Participants had to adopt the director’s perspective in order to successfully select the target in the common ground. For example, as displayed in Figure 2.1, when the director asked for the ‘large balloon’, the item she referred to was the yellow balloon as it was the larger one of the two balloons visible to her. However, the balloon which best-fitted the director’s instruction from the participants’ perspective was the pink balloon. A control trial used an identical grid-image as its corresponding experimental trial, except that the distractor was replaced by an irrelevant object which did not compete with the target referent as a potential referent. Therefore, in the control condition, the employment of the director’s perspective was not a prerequisite for correctly selecting the target referent. For example, the pink balloon presented in the experimental condition was replaced by a broccoli in the control condition, which did not compete with the correct referent—the yellow balloon.

Participants were instructed to perform a 'drag and drop' movement on their selected object using a computer mouse, but consistent with Apperly et al. (2010), the objects did not in fact move on the grid. They were instructed to respond as quickly and accurately as possible. A timeout would occur after 4000ms from the onset of the adjective (or noun for phrases without adjectives), which led to the next instruction or the next grid-image. Accuracy and response time were recorded.

Thirty-two grid-images were grouped into 4 blocks, and each block had 8 grid-images which contained both the experimental and control conditions. Participants were allowed to take breaks between blocks. A grid-image in the control condition was presented at least 8 grid-images apart from its corresponding grid-image in the experimental condition. A practice block containing 2 additional grid images was presented to the participants prior to the 4 test blocks. We created four versions of the experiment by rotating the order of the test blocks.

Upon completion of the director task, participants were asked verbally with three task-related questions: 1. whether they used any strategy to help them do the task and how it helped them; 2 whether they were able to anticipate some of the instructions; 3. what they thought the purpose of the study was. Then participants participated in the OSPAN task.

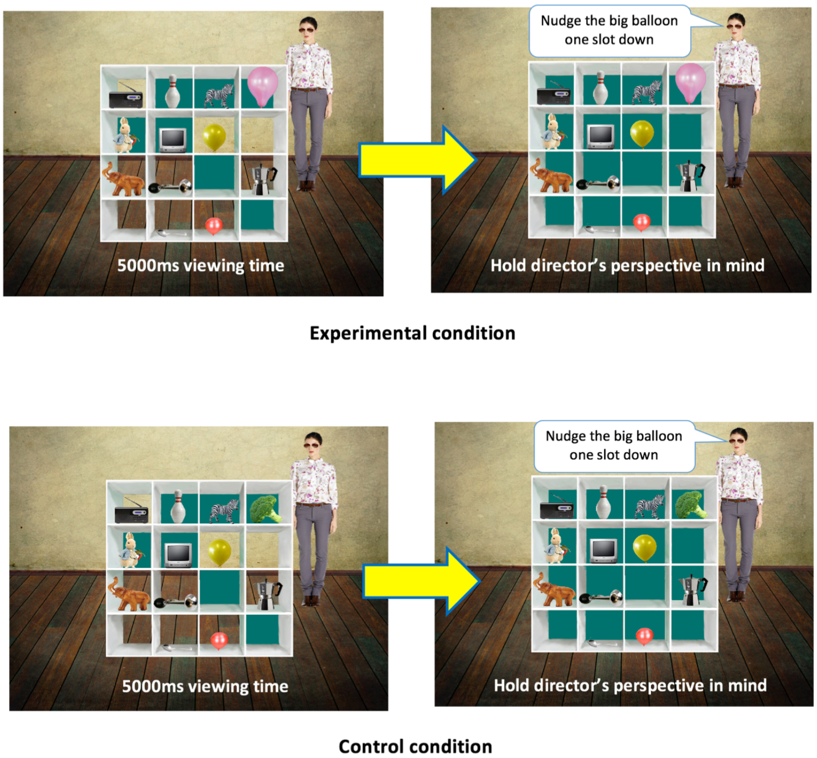


Figure 2.1. Example images for the experimental condition and the control condition of the director task.

**OSPAN task--Working memory capacity**

The present study measured participants' working memory capacity using the OSPAN task (La Pointe & Engle, 1990; Turner & Engle, 1989). The task began with the level of two equation-word pairs, and the level gradually increased from two to five equation-word pairs as the task progressed. Firstly, a mathematical equation (in the form of a division operation followed by an addition operation e.g., (6 / 3) + 4 = 7)) appeared at the centre of the screen, and it could be either a correct or incorrect equation. Participants were asked to verify the equation as quickly and accurately as possible by pressing buttons labelled 'yes' or 'no' on the keyboard for correct and incorrect equations, respectively. Then a word appeared at the centre of the screen, which participants were asked to read out loud. All words were monosyllabic concrete nouns with high frequency. The next equation-word pair was then presented in the same fashion till the last pair. Finally, in the word recall phase, participants were instructed to recall all the words they had just read out loud in the present trial and type them in the correct order using a keyboard. This was considered as a complete trial of the OSPAN task. The OSPAN task consisted of 4 levels ascending from two to five equation-word pairs, and each level consisted of 3 trails. A total number of 12 trials were presented in the same order to each participant.

The word recall for a trial was considered 'perfect' when all of the words presented in that trial were recalled correctly in the correct order. An individual's working memory capacity was calculated by summing the number of words in all perfectly recalled trials. If a trial was not perfectly recalled either because of the wrongly recalled words or the wrongly recalled order, then this trial would not contribute any score to this individual's final working memory capacity score. The range of working memory capacity scores could be from 0 to 42, and higher scores would indicate higher working memory capacities.

## 2.3. Results

### 2.3.1. OSPAN task--Working memory capacity

Participants' working memory capacity scores ranged from 6 (14.3% recall accuracy) to 42 (100% recall accuracy), M = 26.03, SD = 8.89. Higher scores indicate higher working memory capacities (WMC). We were interested in assessing whether participants' performance in the director task varied as a function of their working memory capacity, or as a function of interaction between working memory capacity and other independent variables.

### 2.3.2. Director task--perspective-taking performance

Results from the behavioural data and the eye movement data are reported separately in this section. Only data from critical trials were entered into any of the following analyses. Participants' referent selection was marked as the time point when participants first pressed the left button of the computer mouse to select the object after hearing the critical adjective (e.g., ‘big’). Response to a critical trial was considered as *correct* if participants selected the target object in the common ground. An *egocentric error* referred to the selection of the distractor object in participants’ privileged ground. A *non-egocentric error* referred to the selection of objects or spaces other than the distractor and the target. Participants committed non-egocentric errors on 4.8% of the experimental condition trials, and 3.9% of the control condition trials; response timeouts occurred on 12.5% of the experimental trials, and 10.4% of the control trials. As the causes of the non-egocentric errors were difficult to interpret and were not the central interest of the current study, the non-egocentric errors were excluded prior to the analyses. Critical trials with response timeouts were also excluded prior to the analyses.

#### *2.3.2.1. Behavioural response*

**Egocentric error**

In the control condition, participants made 0% egocentric errors overall. In the experimental condition, individual participants ranged from committing 0% to 100% egocentric errors in a given magnitude condition and from 0% to 67% egocentric errors overall. We also found a large numeric difference in the percentage of trials with percentage egocentric errors across different magnitude conditions: when there were 3 objects in the common ground, participants made an egocentric error on 14.6% of the experimental trials; when there were 5 objects in the common ground, participants made an egocentric error on 22.9% of the experimental trials; when there were 7 objects in the common ground, participants made an egocentric error on 27.7% of the experimental trials; when there were 9 objects in the common ground, participants made an egocentric error on 31.1% of the experimental trials.

To assess whether the possibility of making an egocentric error varied according to the magnitude of common ground and also working memory capacity, we fitted a generalised linear mixed model with the distribution specified as binomial. Due to participants' floor performance in proportion egocentric errors in the control condition, we conducted the following analyses on trials with experimental condition only. The dependent variable was binary, i.e., whether the participant made an egocentric error or not. Matthews, Butcher, Lieven, & Tomasello (2012), a study which examined the effect of object array size and other independent variables on the complexity of sentences children produced in a communication task, used the forward selection procedure to 'explore theoretically motivated models by starting with a simple model with all fixed and random effects, adding two-way interactions in their possible combinations and checking for improvement in fit.' Following the forward selection procedure used by Matthews et al. (2012), we also started with a simple model with only the fixed effects of the Magnitude of Common Ground and Working Memory Capacity (both were entered as continuous variables) and the random effect of Subject (n=37) on intercept[1]. We then added the interaction between these two factors and checked for improvement in model fit. There was no significant improvement but a significant reduction of the goodness-of-fit when we added the two-way interaction of Magnitude x WMC into the model, compared to the model that contained fixed effects of Magnitude and WMC but no interaction, $X^2(1) = 28.944$, $p < .001$. Therefore, the interaction of Magnitude x WMC was not included in the final model.

---

[1] The random intercept of Subject was included in the simple model, as Twisk (2006) states that 'the inclusion of a random intercept is a conceptual necessity for a repeated measures design' (p93). The addition of a random slope of Subject indexed by Magnitude did not lead to the improvement in fit for either of the dependent measure, ps>0.3. Therefore, only random effect of Subject on intercept, but not random slope of Subject indexed by Magnitude, was included into the mixed model for each dependent measure.

In Table 2.1, we reported the parameters of the final model with Egocentric Error as the dependent variable. Results based on the final model revealed that, first of all, participants' working memory capacities significantly predicted their likelihood of committing egocentric errors on experimental condition trials, $F(1, 487) = 15.124$, $p < .001$. With the increase in each unit of WMC score, the odds of committing an egocentric error were multiplied by 0.938 (95% CI [0.908; 0.969]), which suggested that participants with higher working memory were less likely to make egocentric errors on experimental condition trials. Independent of this WMC effect, we also found that the magnitude of common ground significantly predicted the likelihood of participants making egocentric errors, $F(1, 487) = 8.523$ , $p = .004$. With each unit of increase in the magnitude of common ground, the odds of committing an egocentric error were multiplied by 1.158 (95% CI [1.049; 1.278]), which suggested that participants were more likely to commit egocentric errors when the magnitude of common ground was larger. The interaction between the effects of WMC and magnitude of common ground was not included into the final model, which suggested that working memory capacity influenced participants' egocentrism in this director task in general, but it did not modulate the effect of magnitude on participants' egocentric error rate.

Table 2.1. Fixed effects in the generalized linear mixed model with Egocentric Error (binary) as the dependent variable.

| | | | | | Confidence Interval | | | Confidence Interval for Exp(β) | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | β | S. E. | t | Significance | Lower | Upper | Exp(B) | Lower | Upper |
| Intercept | -.509 | .5190 | -.981 | .327 | -1.529 | .511 | .601 | .217 | 1.666 |
| WMC | -.064 | .0164 | -3.889 | .000 | -.096 | -.032 | .938 | .908 | .969 |
| Magnitude | .147 | .0502 | 2.919 | .004 | .048 | .245 | 1.158 | 1.049 | 1.278 |

**Response time**

Participants' response time to a critical instruction was measured from the onset of the critical adjective word (i.e., the onset of the integration phase) to the point where they selected a referent by pressing down the left button of the computer mouse. Only the critical trials on which participants responded correctly were included into the response time (RT) analysis, resulting in a further 10.3% of all the critical trials excluded prior to subsequent analyses. Only one response time data point (1268ms) fell 2.5 standard deviations beyond the mean RT, and this was removed prior to the RT analysis.

A linear mixed model was fitted with response time as the dependent variable. Similar to the egocentric error data, we also started with a simple model with the fixed effects of the Magnitude of Common Ground, Working Memory Capacity and ExpCondition, and with the random effect of Subject (n = 37) on intercept. The Magnitude of Common Ground and Working Memory capacity were both entered as continuous variables, and ExpCondition was entered as a binary variable (1 = Experimental condition, 0 = Control condition). Based on the simple model, we then added the interactions between these three variables in all possible combinations and checking for improvement in fit. There was no significant improvement when we added any of the two-way or 3-way interactions, compared to the model whose fixed effects contained only Magnitude, WMC and Experimental/control condition without any interaction, all $X^2$ (1)s < 0.825, *p*s > .364. Therefore, no 2-way or 3-way interaction was included in the final model.

In Table 2.2, we reported the parameters of the final model with Response Time as the dependent variable. Results based on the final model revealed that only the magnitude of common ground significantly predicted participants' response time errors, $F(1, 846.230) = 40.067$, $p < .001$. With one unit of increase in the magnitude of common ground, participant's response time increased 34.781 (95% CI [23.996; 45.566]). The ExpCondition and WMC factors did not predict response time data, $p$ =.588 and $p$ = .377 respectively. Overall, the response time results suggest that the more objects presented in the common ground, the longer it took for participants to select the target, regardless of whether there was any distractor competing against the target as a potential referent, and regardless of participants' working memory capacities.

Table 2.2. Fixed effects in the linear mixed model with Response Time (ms) as the dependent variable.

| | | Confidence Interval | | | | |
|---|---|---|---|---|---|---|
| Variable | β | Lower | Upper | S. E. | t | Significance |
| Intercept | 3002.117 | 2782.461 | 3221.774 | 109.140 | 27.507 | .000 |
| [ExpCondition=0] | 13.691 | -35.912 | 63.294 | 25.272 | .542 | .588 |
| [ExpCondition=1] | $0^b$ | 0 | . | . | . | . |
| WMC | -3.348 | -10.932 | 4.236 | 3.740 | -.895 | .377 |
| Magnitude | 34.781 | 23.996 | 45.566 | 5.495 | 6.330 | .000 |

Notes: ExpCondition (Experimental condition = 1, control condition = 0). The experimental condition was specified as the reference category in the mixed model.

#### *2.3.2.2. Eye movements*

The same trials used for the response time analysis were used in eye movement analyses. Each critical trial was divided into two time windows of interest, namely the *perspective encoding phase* and *perspective integration phase* (see Figure 2.2). The participants' eye movements from these two phases were analysed and reported separately. The initial 5000ms from the onset of a new grid-image was treated as the *perspective encoding phase*. During this time, participants did not receive any audio input and they were instructed to examine and remember what the director could and could not see. This was the only phase when perspective encoding could take place, as the visual cues indicating an object's ground status would be removed after the initial 5000ms. The *perspective integration* *phase* was defined as the period from the onset of the adjective of the critical instruction (e.g., 'big' in 'nudge the big balloon one slot up') until participants' referent selection. Participants' referent selection was marked as the time point when participants first pressed the left button of the computer mouse to select an object. During the perspective integration phase, participants were expected to integrate the encoded perspective information with the critical linguistic information (i.e., the scalar adjective in the critical instruction) to successfully resolve reference. Note that there was an additional time period during the critical instructions, starting from the onset of the first word of the instruction and ending prior to the critical scalar adjective, i.e., 'Nudge the...'. However, given that this phase only lasted an average of just 914ms and the eye movement data from this phase were likely to be noisy due to participants attending to the onset of the instruction, the eye movements from this time window were not analysed.

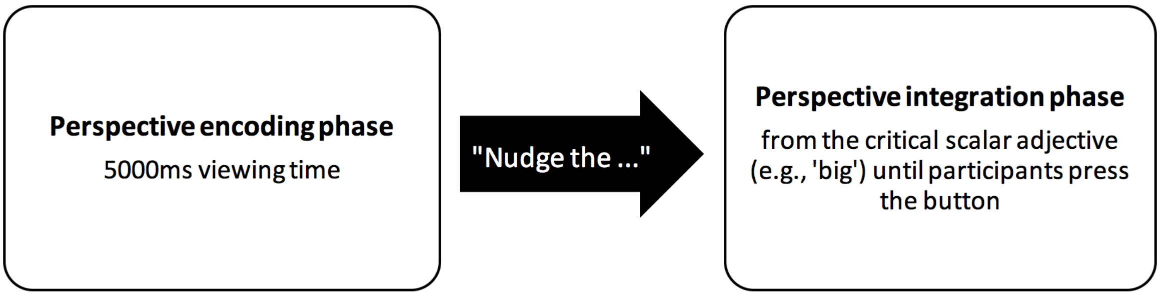


Figure 2.2. Illustrations of the structure of perspective encoding phase and perspective integration phase of a critical trial.

During the *perspective encoding phase* the critical information about the target (i.e. the adjective and the noun) had not yet been provided to participants. As the only difference between the grid-image used in the experimental and its corresponding control condition was that a distractor in the experimental condition was replaced by an irrelevant object in the control condition, we expected participants in both conditions to examine the two grid-images similarly prior to the onset of the critical adjective. However, during the *perspective integration phase*, we expected participants to behave differently when interpreting the instructions from the experimental condition versus the control condition. The reason for this expectation was that in the experimental condition, participants could only identify the correct referent by taking the director's perspective into account; while in the control condition the same correct referent could be identified without perspective integration.

Only fixations started 200ms after the onset of one phase and before the offset of the same phase were counted as fixations during that phase. Planning a saccade takes 200ms in general and thus fixations in the first 200ms are unlikely to reflect participants' examination of the experimental stimuli. Fixations shorter than 100ms were excluded from the analyses, as they were more likely to be a pause in a sequence of saccades, rather than an actual fixation. For a fixation that lasted beyond the offset of its phase, only the portion from the beginning of the fixation until the offset of the phase was calculated and included in the analysis.

Two dependent variables were calculated from the eye movement data, one from each time window, based on both findings from previous studies and also our research hypotheses. For the encoding phase, based on the previous findings that participants showed preference/anticipation effect to common ground objects than to privileged ground objects (Barr, 2008; Heller et al., 2008), we expect participants to dwell and fixate more on common ground objects. Common ground was therefore the focus of our analysis of the perspective encoding phase, as objects in the common ground form the set of potential referents if participants were successfully taking the director's perspective into account. Three steps were taken to compute a *Common Ground Preference Score* for each magnitude. First, we calculated each participant's expected dwell time on common ground items on the null hypothesis that they were equally likely to fixate on each of the 12 objects in the array. To do so we summed participants' total dwell time on common ground and privileged ground objects for each trial, and multiplied this by the fraction of common ground items (e.g., for the 3-common-ground condition, this fraction would be 3/12). Next we subtracted this expected dwell time from the observed dwell time on common ground items. Finally, we divided this remainder by the number of common ground items. This procedure yielded a preference score for common ground items, expressed in milliseconds, the value of which should be zero on the null hypothesis that participants had no preference for common ground items.

During the perspective integration phase, people tend to fixate on an object prior to selection. Participants' final fixation on the target before they correctly selected the target was interpreted as the marker of the end of participants' decision-making process (Keysar et al., 2000). The latency to final target fixation was thus calculated by many studies employing this particular version of the director task to measure the amount of time participants took to resolve a referential problem and make a decisive fixation (e.g., Keysar et al., 2000; Lin et al., 2010; Wang et al., submitted; Wu & Keysar, 2007). A longer decision-making process was expected to reflect more difficulty in resolving the reference. While some other eye-tracking studies have also calculated participants' proportion of fixations to the target and/or to the distractor, and examined how participants' target advantage scores (proportion of target fixations minus proportion of competitor fixations) change over time during perspective integration phase (e.g., Barr, 2008; Brown-Schmidt, 2009b), this measure was not appropriate for the current design. The aforementioned studies all used a variant of the director task which contained instructions with temporary ambiguity, and the referring expressions could be disambiguated either at an early point or at a late point of the instructions. For example, Barr (2008) used objects that were phonological competitors (e.g., a bucket and a buckle), and thus made the target of the speaker's referring expression temporarily ambiguous until they heard the last part of the object name (e.g., ‘et’ or ‘le’). Brown-Schmidt (2009b) used images and instructions of ‘the *cow* that’s wearing shoes’ versus ‘the *cow* that’s wearing glasses’. By comparing participants’ fixations to the targets and/or distractors between conditions with either early or late point of disambiguation, researchers were able to investigate the more fine-grained time course of the effect of common ground during referential communication. However, the version of the director task used in the present study was not designed with such temporary ambiguity, but was designed with scalars (e.g., big/small) with the aim of maximising the chances of committing an egocentric error. In the current version of the director task, the distractors were only present in the experimental condition but not in the control condition. Therefore, the current design does not allow for meaningful comparison between the experimental condition and control condition on proportion of fixations to target / distractor were not theoretically. Additional analyses were conducted to examine the effect of common ground size and working memory capacity on participants’ proportion of fixations on the distractor during the experimental condition only, but no significant effect was found.

**Perspective encoding phase: common ground preference**

Firstly, one-sample t-tests were conducted to test whether participants’ aggregated common ground preference scores in each magnitude condition differed from zero. A common ground preference score significantly higher than zero would indicate a preference for common ground, and a score significantly lower than zero would indicate a preference for privileged ground. For 3 and 5 magnitude condition, the common ground preference scores were both significantly larger than zero, $t(36) = 7.650$, $p < .001$ and $t(36) = 5.087$, $p < .001$, respectively, suggesting a viewing preference for common ground object. For 7 magnitude condition, no significant preference for either common or privileged ground objects was observed, $t(36) = -.166$, $p = .869$. Interestingly, for 9 magnitude condition, the common ground preference score was significantly lower than zero, $t(36) = -5.105$, $p < .001$, suggesting a viewing preference for privileged ground objects.

Secondly, a linear mixed model was fitted with participants' Common Ground Preference Scores as the dependent variable. We started with a simple model with the fixed effects of Magnitude, WMC and ExpCondition, and with the random effect of Subject on intercept. We then added the interactions between these three variables in all possible combinations and checking for improvement in fit. Only the addition of the interaction of WMC x Magnitude significantly improved the goodness-of-fit of the model, $X^2(1) = 6.026$, $p = .014$. There was no further improvement to the goodness-of-fit of the model when other two-way or 3-way interactions were added. Therefore, the final model included the fixed effects of Magnitude, WMC, ExpCondition and WMC x Magnitude, and the random effect of Subject on intercept.

In Table 2.3, we reported the parameters of the final model with common ground preference scores as the dependent variable. Results based on the final model found that, first of all, participants' working memory capacities significantly predicted their common ground preference scores, $F(1, 133.446) = 6.319$, $p = .013$. With the increase in each unit of WMC score, participants' common ground preference score increased 4.768 units (95% CI [1.017; 8.520]), which suggested that participants with higher working memory were more likely to dwell longer at each common ground object compared with privileged ground objects. The magnitude of common ground also significantly predicted participants' common ground preference scores, $F(1, 831.526) = 14.039$, $p < .001$. With one unit of increase in the magnitude of common ground, participant's common ground preference score decreased 23.171 units (95% CI [-35.310; -11.033]). This suggested that participants were more likely to dwell longer at each common ground object when there were fewer objects in the common ground. The ExpCondition did not predict common ground preference scores, $F(1, 833.089) = .130$, $p = .719$, which was consistent with our expectation.

Moreover, a significant interaction of WMC with Magnitude was found, $F(1, 831.262) = .6.046$, $p = .014$. To follow this interaction up, firstly, we fitted a linear mixed model on participants' common ground preference scores for each magnitude group, with the fixed effect of WMC and the random effect of Subject on intercept. For the 3-common-ground magnitude condition, WMC only marginally significantly predicted participants' common ground preference scores, $p = .095$, $\beta = 4.129$. For 5, 7 and 9-common-ground magnitude condition, WMC was not a significant predictor at all, $p$s > .600. Secondly, we divided participants into 3 groups based on their WMC scores, i.e., low WMC group (n=12, WMC <= 20), medium WMC group (n=13, 21<=WMC<=29) and high WMC group (n=12, WMC>=30), and fitted a linear mixed models on participants' common ground preference scores for each WMC group. The magnitude of common ground significantly predicted common ground preference scores for each WMC group, $p$s < .001. With one unit of increase in the magnitude of common ground, participant's common ground preference score decreased 26 units (95% CI [-33.665; -19.166]) for low WMC group, 44 units (95% CI [-50.453; -37.431]) for medium WMC group, and 41 units (95% CI [-46.755; -34.377]) for high WMC group respectively. This seemed to suggest that all participants fixated longer onto each privileged ground object over each common ground object when there were fewer objects in the privileged ground, and fixated longer onto each common ground object over each privileged ground object when there were fewer objects in the common ground. However, compared to participants with low working memory capacities, participants with medium to high working memory capacities revealed a greater degree of sensitivity to the experimental manipulation of the relative size of common ground versus privileged ground.

Table 2.3. Fixed effects in the linear mixed model with Common Ground Preference Scores (ms) as the dependent variable.

| Variable | β | Confidence Interval Lower | Upper | S. E. | t | Significance |
|---|---|---|---|---|---|---|
| Intercept | 149.440 | 45.118 | 253.763 | 52.766 | 2.832 | .005 |
| [ExpCondition=0] | 3.221 | -14.342 | 20.783 | 8.948 | .360 | .719 |
| [ExpCondition=1] | 0[b] | 0 | . | . | . | . |
| WMC | 4.768 | 1.017 | 8.520 | 1.897 | 2.514 | .013 |
| Magnitude | -23.171 | -35.310 | -11.033 | 6.184 | -3.747 | .000 |
| WMC x Magnitude | -.547 | -.983 | -.110 | .222 | -2.459 | .014 |

**Perspective integration phase: latency to final target fixation**

A linear mixed model was fitted on the latency to final target fixation data. Subject (n=37) was added as a random effect on intercept. We started with a simple model with the fixed effects of Magnitude, WMC and ExpCondition, and with the random effect of Subject on intercept. Based on the simple model, we then added the interactions between these three variables in all possible combinations and checking for improvement in fit. There was no significant improvement when we added any of the two-way or 3-way interactions, compared to the model whose fixed effects contained only Magnitude, WMC and ExpCondition without any interaction. Therefore, no 2-way or 3-way interaction was not included in the final model.

In Table 2.4, we reported the parameters of the final model with Latency to Final Target Fixation (ms) as the dependent variable. Similar to the response time data, results based on the final model revealed that only the magnitude of common ground significantly predicted participants' response time errors, $F(1, 770.589) = 26.879$, $p < .001$. With one unit's increase in the magnitude of common ground, participant's response time increased 43.728 (95% CI [27.171; 60.285]). The ExpCondition and WMC factors did not predict response time data, $p$ =.588 and $p$ = .377 respectively. Overall, the more objects presented in the common ground, the longer it took for participants to launch a final fixation on the target before making a correct response, regardless of whether there was any distractor competing against the target as a potential referent, and regardless of participants' working memory capacities.

Table 2.4. Fixed effects in the linear mixed model with Latency to Final Target Fixation (ms) as the dependent variable.

| | | Confidence Interval | | | | |
|---|---|---|---|---|---|---|
| Variable | β | Lower | Upper | S. E. | t | Significance |
| Intercept | 2488.623 | 2203.555 | 2773.691 | 142.085 | 17.515 | .000 |
| [ExpCondition=0] | 17.301 | -57.384 | 91.988 | 38.046 | .455 | .649 |
| [ExpCondition=1] | 0[b] | 0 | . | . | . | . |
| WMC | -6.482 | -16.096 | 3.133 | 4.746 | -1.366 | .180 |
| Magnitude | 43.728 | 27.171 | 60.285 | 8.434 | 5.185 | .000 |

#### *2.3.2.3. Supplementary analyses on participants' reported encoding strategy*

During the debriefing session, 20 out of 37 participants explicitly reported that they paid more attention to common ground objects when there were fewer objects in the common ground than in the privileged ground, and switched their attentional preference to privileged ground objects when there were fewer objects in the privileged ground during the 5s viewing phase. This 'switching' strategy was the most common strategy participants reported during the director task. We therefore divided participants into two groups based on their report of this switching strategy: those 20 participants who had explicitly reported their use of this specific strategy was in the 'strategy' group, the remaining 17 participants who did not report the use of this strategy were in the 'no-strategy' group. While we had not planned to include this reported strategy use as a factor in our main analyses, we did check its effects on each dependent variable.

As over half of the participants explicitly reported that they used a 'switching' strategy during the perspective encoding phase, a 2x4 repeated-measure ANOVA was conducted on participants' common ground preference scores with the reported strategy and the magnitude of common ground as two factors, in order to check whether the inclusion of this strategy factor exerted any effect. The common ground preferences scores were collapsed between experimental and control condition, as no difference was expected to be found between the experimental and control condition, and this expectation was also confirmed by the non-significant effect of experimental/control condition observed from the linear mixed model on common ground preferences scores. The relevant descriptive data was reported in Table 2.4. Results from ANOVA showed that there was a significant main effect of magnitude, $F(3, 105) = 77.803$, $p < .001$, $\eta p2 = .690$, but no main effect of reported strategy, $F(1, 35) = 21.227$. $p = .365$, $\eta p2 = .378$. Interestingly, a significant interaction of magnitude x strategy was observed, $F(3, 105) = 8.712$, $p = .001$, $\eta p2 = .199$. To further explore this significant interaction, independent t-tests were conducted between 'strategy' group and 'no-strategy' group for each array size. In the 3 magnitude condition, strategy group showed significantly higher degree of common ground preference than the no-strategy group, $t(35) = -2.420$, $p = .021$. In the 5 or 7 magnitude condition, no difference was found between the groups, $t(35) = -.966$, $p = .360$ and $t(35) = 1.072$, $p = .304$, respectively. Conversely, in the 9 magnitude condition, strategy group showed significantly higher degree of preference towards privileged ground than the no-strategy group, $t(35) = 2.052$, $p = .048$. Results suggest that participants who reported use of 'switching' strategy showed a neat correspondence between their eye movements and the strategy reported. While both groups showed the same qualitative pattern, the eye movements of participants who reported a 'switching' strategy were affected by the size of common ground to a greater degree than those in the no-strategy group.

Table 2.4. Means and standard errors of common ground preference score for 'strategy' and 'no-strategy' groups respectively.

| | Strategy (N=20) | | | | | | No-strategy (N=17) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total dwell time on common ground (ms) | | Total dwell time on privileged ground (ms) | | Common ground preference (ms) | | Total dwell time on common ground (ms) | | Total dwell time on privileged ground (ms) | | Common ground preference (ms) | |
| magnitude | Mean | SE | Mean | SE | Mean | SE | Mean | SE | Mean | SE | Mean | SE |
| 3-common-ground | 1574 | 80 | 1700 | 68 | 249 | 24 | 1209 | 122 | 2036 | 174 | 133 | 44 |
| 5-common-ground | 1883 | 84 | 1450 | 101 | 99 | 17 | 1720 | 154 | 1607 | 167 | 67 | 30 |
| 7-common-ground | 1943 | 91 | 1541 | 98 | -13 | 13 | 2049 | 123 | 1331 | 143 | 11 | 19 |
| 9-common-ground | 1965 | 86 | 1334 | 104 | -57 | 11 | 2253 | 114 | 1043 | 114 | -24 | 12 |

Similarly, a 2x4 repeated-measure ANOVA was conducted on participants' percentage egocentric errors in experimental condition, with reported strategy use and magnitude of common ground as two factors. The analysis revealed no significant main effect of strategy use ($p$ = .815) or any interactions involving this factor ($p$ = .260). A 2x4x2 ANOVA was then carried out on response time, including reported strategy, magnitude of common ground, and condition as factors. The analysis revealed no significant main effect of reported strategy ($p$ = .530) or any interactions involving this factor ($p$ = .957 for condition x strategy, p = .248 for magnitude x strategy, $p$ = .828 for condition x magnitude x strategy). A 2x4x2 ANOVA was also carried out on latency to final target fixation, including reported strategy use, magnitude of common ground, and condition as factors. This analysis revealed no significant main effect of this strategy use ($p$ = .151) or any interactions involving this factor ($p$ = .697 for condition x strategy, $p$ = .410 for magnitude x strategy, and $p$ = .666 for condition x magnitude x strategy).

In brief, only participants' eye movements in the perspective encoding phase were influenced by this reported switching strategy, while other dependent variables were not affected.

## 2.4. Discussion

Existing research has led researchers to strikingly inconsistent conclusions about the relationship between working memory and the use of perspective during communication. Some studies found that individual differences in working memory capacity affected perspective use in communication (Lin et al., 2010; Wardlow, 2013); whereas other studies failed to observe any effect of individuals' working memory capacities on perspective taking (Cane et al., 2016; Nilsen & Graham, 2009). To cast new light on this important relationship we followed previous studies by measuring participants' working memory capacities, but in addition we adapted the commonly-used director task so that participants were forced to remember the director's perspective, and for the first time we manipulated memory load by systematically altering the relative size of common ground versus privileged ground. With our adapted director task we succeeded in finding distinct memory effects during perspective encoding, and during the integration of perspective information with the director's instructions. Fixation patterns during encoding were sensitive to the number of items in the common versus privileged ground, but were not affected by participants' working memory capacities. Egocentric errors during integration were sensitive both to the number of items in the common ground and to participants' working memory capacities, but these effects were statistically independent, suggesting that they arose from distinct processes. We expand our discussion of these results below.

***Perspective encoding.*** During the encoding phase participants' viewing preference towards common ground was the focus of our analysis. Our study showed that the size of common ground had a significant linear effect on participants' eye movements. Interestingly, when common ground size was greater, rather than devoting more resources to encoding relevant common ground information as we predicted, participants instead paid more visual attention to each privileged ground object compared to each common ground object. This finding might appear to be at odds with the 'anticipatory effect' found by Barr (2008) that a preference for common ground items was typically observed prior to the point when participants were required to integrate information about a speaker's perspective with her instructions. However, two points are noteworthy. First, in studies reporting anticipatory effects, the number of objects in the common ground was typically low (e.g., 3 common ground objects and 1 privileged ground object in Barr (2008)), and within the range where we too found preferential attention to common ground items. Second, in previous studies participants was not required to remember which items are in the common versus privileged ground, because this information remained available at the point of need. We speculate that, under the enforced memory demands of our adapted director task, participants might have encoded information in the array strategically, with a preference for remembering the smaller set of items, whether they were in the common or privileged ground.

Supportive of our speculation, 20 out of 37 participants in the debriefing session explicitly confirmed their employment of the strategy of encoding whichever was the smaller set of items. Our supplementary analyses suggested that the group who reported their use of strategy were indeed doing what they reported, as their common ground preference scores were significantly higher than those of the no-strategy group when the number of items in the common ground was smaller, and significantly lower than those of the no-strategy group when the number of items in the common ground was larger. This strategy may be task-specific, or may reflect the flexibility with which people approach the problem of tracking common ground in discourse. Outside of the laboratory people surely do sometimes identify privileged ground items specifically for the purpose of not incorporating them into discourse, as when reminding oneself not to mention the salient but secret surprise party being held for our interlocutor. Either way, it is noteworthy that the strategic effects observed during encoding did not have any observed consequences for success at actually using perspective information during the integration phase. This result accords with a study on language production which suggests that people do not only have difficulty avoiding leaking privileged knowledge, but that reminding oneself not to mention the privileged information makes it actually harder to guard against giving the secret away (Lane, Groisman, & Ferreira, 2006).

Apart from the effect of the size of common ground, we also found a significant linear effect of working memory capacity during perspective encoding. Participants with higher working memory capacities were more likely to dwell longer at each common ground object compared with privileged ground objects during perspective encoding phase, compared to participants with lower working memory capacities. However, when we followed up on the significant interaction of Magnitude x WMC, the effect of WMC was only marginally significant for the 3-common ground magnitude condition, but not for other greater magnitude conditions. Moreover, follow-up analyses on the significant interaction of Magnitude x WMC showed that participants with medium and high working memory capacities revealed a greater degree of sensitivity to the experimental manipulation of the relative size of common ground versus privileged ground, compared to participants with low working memory capacities. That is, participants with medium and high working memory capacities were more likely to fixate longer onto each privileged ground object over each common ground object when there were fewer objects in the privileged ground, and were more likely to fixate longer onto each common ground object over each privileged ground object when there were fewer objects in the common ground. This is the opposite pattern to what would be expected if greater working memory capacity enabled participants to focus more reliably on common ground items. It is also noting that the eye movements during perspective encoding only indicated how participants distributed their attention to items in the common ground versus privileged ground. It was therefore unclear, both in principle and in terms of the actual pattern of results, how this attention allocation pattern during online perspective encoding phase contributed to participants' final performance on the director task which was revealed by the three dependant variables (i.e., egocentric errors, response time and latency to final target fixations) in the perspective integration phase.

***Perspective integration.*** During this phase, participants had to integrate the director's instruction with information about her perspective. Our principal interest was in the 'egocentric' effects whereby responses were slower or more error prone in the experimental condition where a response based only on participants' own perspective would differ from a response that took account of the director's perspective, compared with the control condition where responses based on either privileged ground or common ground would not conflict with each other. In the present study, this 'egocentric' effect was only observed in the proportion egocentric errors, and not in response times or latencies to final fixation, which is consistent with some previous studies (Apperly et al., 2010; Dumontheil et al., 2010).

Firstly, participants in the present study made more egocentric errors as the size of common ground increased. Increasing the number of items in the common ground also resulted in slower responses and longer latencies to final target fixation, but unlike the case of errors, experimental and control trials were equally affected. Secondly, we observed a significant effect of working memory capacity exclusively on participants' proportion egocentric errors. Individuals with higher working memory capacities tended to commit fewer egocentric errors in the experimental condition where common ground information needed to be integrated with the speaker's message. This finding is consistent with Lin et al. (2010)'s study where participants with lower working memory capacities showed delayed response time to select the target referent; however, it is inconsistent with Cane et al. (2016) and Nilsen and Graham (2009) where correlations between participants' working memory capacity scores and their perspective-taking performance failed to emerge. We think it is likely that our study greatly enhanced the chances of detecting memory effects because, unlike previous studies, it guaranteed a demand of having to remember the speaker's perspective, rather than being able to compute it at the point of need. However, this would still leave the positive results of Lin et al. (2010) without an explanation, as Lin et al.' s study did not require participants to retain information about perspectives. In this case, it seems plausible that Lin et al.'s analysis strategy of only selecting participants whose working memory capacity scores were within the top 20 or bottom 20 percentiles of their pre-test sample was likely to have enhanced their chances of detecting a significant effect of working memory capacity.

Critically, while both common ground size and working memory capacity influenced the number of egocentric errors participants made on experimental condition trials, these factors did not interact. That is to say, the number of egocentric errors participants with lower working memory capacities committed was no more influenced by common ground size than for participants with higher working memory capacities. This lack of interaction between the magnitude of common ground and working memory capacity could potentially provide some informative insights into the memory processes involved in perspective taking during communication. In the present study, an egocentric error could be made due to two reasons:1) participants had difficulty in successfully remembering and holding in mind relevant perspective information of what was in common ground; 2) participants had difficulty in successfully integrating the correctly encoded common ground information with the verbal messages from the director. If both the magnitude of common ground and working memory capacity had affected the number of egocentric errors participants committed by affecting the likelihood of successfully remembering and holding in mind relevant perspective information, then we would have observed an interaction of Magnitude x WMC. Similarly, if both magnitude and working memory capacity had affected the likelihood of successfully integrating the relevant perspective information with linguistic input then we would have also observed an interaction of Magnitude x WMC. However, our results provided no evidence for either of these two accounts.

The most plausible explanation was that the size of common ground and working memory capacity might have affected participants' perspective-taking performance via different routes. Firstly, we propose that participants' working memory capacities were likely to have influenced the number of egocentric errors they committed by directly affecting the likelihood of successfully integrating the critical constraint from information about the director's perspective with the linguistic information from the director's instruction. This speculation is consistent with common views about working memory as an executive-attention control mechanism that helps integrate diverse sources of information together with goal-relevant information over short timescales (e.g., Conway, Kane, & Engle, 2003). The perspective integration process is likely to depend heavily upon one's working memory capacities. The lower one's working memory capacity, the fewer memory resources one could utilize to integrate perspective information into communication, and the more egocentric errors one would commit.

In contrast, the size of common ground may have affected adults' perspective use mainly via a longer term memory route. we propose that the size of common ground was likely to have affected the number of egocentric errors participants committed by affecting the likelihood of their being able to successfully encode and retain what was in common ground. Recall that the current results found that as the size of common ground increased, participants generally devoted progressively less attention to common ground items. We suggest that one plausible consequence of this effect was that when common ground was larger, participants retained a less accurate record of what was in common ground in their longer term memory, and thus were less likely to correctly reject the distract in the privileged ground and choose the target object in the common ground. The size of common ground was likely to have affected the rates of egocentric errors through such a long-term memory route. This speculation is consistent with our finding that the effect of common ground size on egocentrism was insensitive to participants' working memory capacities. The demand to hold in mind an increasing amount of common ground information served as a separate contributing factor to the egocentric effect observed in the current study. This detrimental effect of an increasing common ground size could not be simply compensated by having more available working memory resources. Importantly, this effect is also consistent with other studies in the literature suggesting that constraints on discourse processing operating over time scales of more than a few seconds depend upon retention and retrieval of information from long-term memory (Ericsson & Kintsch, 1995; Graesser et al., 1997).

Some researchers might argue that the observed effect of working memory on egocentrism the current study could be alternatively explained by participants' motivation to succeed on both the director task and the OSPAN task. Admittedly, it was possible that participants who were more motivated to succeed on the OSPAN task might also be more motivated to succeed on the OSPAN task. However, given that OSPAN task is a cognitively demanding task and is generally believed as a reliable and valid measure of working memory capacity (e.g., Unsworth & Engle, 2007; Unsworth, Redick, Heitz, Broadway, & Engle, 2009), it is unlikely that participants' performance on OSPAN task was a mere reflection of their motivation in succeeding on the task. Participants who were less motivated to succeed on the OSPAN task might get a lower working memory capacity score, even when they had the working memory capacity required by the task. On the contrary, participants with high motivation in succeeding on the OSPAN task would still be constrained by the cognitively demanding nature of the OSPAN task if their working memory capacity was rather limited. Nonetheless, future studies could include a measure which is not expected to correlate with participants' performance in either the director task or the OSPAN task to rule out this motivational explanation.

To conclude, what was novel in the present study was that we adapted a widely-used director task to capture the memory demand of holding in mind someone's perspective, which is commonly embedded in our everyday communication but was overlooked in the previous literature. Apart from this, we systematically varied the size of common ground in the adapted version of director task, and also measured participants' working memory capacities. Our findings are consistent with suggestions that perspective integration draws heavily upon domain-general working memory resources. Participants exhibited higher degree of egocentrism when common ground size was larger, and when ones' working memory capacities were smaller; but these two factors did not interact. We propose that such results provide some evidence for dissociable effects of long-term memory and working memory on perspective encoding and perspective integration respectively. It's important for future studies to dissociate the sub-processes involved in taking other people's perspectives, and to incorporate into laboratory communication tasks the varying memory loads interlocutors may experience in everyday communication. This will help specify the mechanisms underlying how different memory processes contribute to different sub-process of perspective-taking.

CHAPTER III

# The cognitive demands of remembering a speaker's perspective and managing common ground size modulate 8- and 10-year-olds' perspective-taking abilities

## 3.1. Introduction

Perspective taking can be seen as a common instance of 'theory of mind' use. Until relatively recently, a large number of developmental studies on the research area of 'theory of mind' focused on answering the question of when children develop this understanding of others' mental states and perspectives (Astington & Gopnik, 1991; Perner et al., 1987). Classic accounts suggest that children are egocentric, and incapable of understanding others' perspectives before around 7 years (Piaget, 1959; Piaget & Inhelder, 1956). More recent evidence suggests that 'theory of mind' concepts develop significantly earlier, between 2 and 7 years (e.g., Wellman, Cross, & Watson, 2001), and may even be present in infants (e.g., Kovács, Téglás, & Endress, 2010; Moll & Tomasello, 2006; Onishi & Baillargeon, 2005; Sodian, Thoermer, & Metz, 2007). In contrast, recent research on older children and adults, who have clearly developed a basic understanding of mental states, shows that they are still egocentrically biased especially when facing communicative situations where differing perspectives have to be taken into account (e.g., Dumontheil et al., 2010; Keysar et al., 2000, 2003). This suggests that employing theory of mind abilities in communication requires much more than just acquiring the necessary theory of mind concepts.

As introduced in Chapters 1 and 2, a large number of studies employing this director task on healthy adults have demonstrated that adults suffer from interference of their own privileged perspective (Apperly et al., 2010; Keysar et al., 2000, 2003; Wu & Keysar, 2007). Children, who are less experienced communicators than adults, may suffer more from this egocentric bias when required to accommodate differing perspective in the director task. Although eye movement data from Nadig and Sedivy (2002) suggested that even 6-year-old children were sensitive to their interlocutor's limited perspective and were able to use this perspective information from the early stage of language processing, egocentrism did not completely evaporate in their study. Indeed, the first experiment in Nadig and Sedivy (2002) suggested that children speakers were not as reliable as adult speakers at providing adjectives to fully disambiguate the referents when there are two similar candidates in the common ground. Epley, Morewedge, and Keysar (2004) demonstrated that 4- to 12-year-old children acted more egocentrically in the task (as they reached for hidden referents more frequently) compared to the adult participants. They provided a plausible explanation that adults and children appeared to not differ in the initial processing stage during which they both processed information egocentrically, but differed in the later adjustment stage during which adults were more capable to correct their egocentric bias and accommodate their interlocutor's differing perspective than children. Moreover, Epley et al. (2004) also observed an incremental improvement in performance as children's age increased. Convergingly, Dumontheil et al. (2010) and Symeonidou, Dumontheil, Chow, and Breheny (2016) found that children's ability to accommodate a speaker's limited perspective continued to develop through childhood to adolescence till adulthood. Similarly, Frick, Möhring, and Newcombe (2014) used an experimental paradigm where two agents saw the same sets of objects from two different visuo-spatial perspectives from which participants' own perspective could also differ. This study found that the capacity to inhibit egocentric choices and to recognize what agents could see from their own perspectives gradually developed between 5 to 8 years of age. These findings suggest that the use of information about others' perspectives goes through prolonged developmental improvements even after the acquisition of the basic theory of mind competence.

In the present study, we first sought to identify the cognitive factors, especially memory factors, that might affect children's abilities to infer and utilize a speakers' perspective during language comprehension. Secondly, we considered whether the same factors would drive age-related improvements in children's perspective-taking performance. As reviewed in Chapter 2, most studies examining the relationship between memory capacities and perspective-taking performance in referential communication have been conducted on adults, but the findings are currently mixed (e.g., Brown-Schmidt, 2009b; Cane et al., 2016; Lin et al., 2010). Little research has been done to directly examine the memory factors that affect children's ability to use a speaker's perspective in referential communication tasks e.g., the director task. The most relevant developmental evidence came from Nilsen and Graham (2009), where they found a positive correlation in children aged three to five between their inhibitory control skills and their perspective-taking abilities in language comprehension. A similar tendency was also observed for the measure of working memory capacities, but it did not reach significance. A more recent study conducted by Wang, Ali, Frisson, and Apperly (2016) observed that 10-year-olds performed less egocentrically than 8-year-olds in the director task overall, and both age group committed more egocentric errors when a director's instruction sentence was complex (e.g., 'nudge the large jar one slot up') rather than simple (e.g., 'nudge the large jar'). However, both 8- and 10-year-olds were affected by the cognitive demands of integrating complex messages with the speaker's limited perspective to the same degree, suggesting the age-related improvements in children's perspective-taking performance was not driven by the development of executive capacities that might be necessary for meeting this demand. Some other studies, though not directly employing the director task, have also demonstrated the important role of memory in children's performance in referential communication. For example, a previous study by Dahlgren & Sandberg (2008) found that free recall capacities measured by both verbal and object free recall tasks contributed significantly to performance in a referential communication task for both children with autism spectrum disorder and normally developing children. Another study found that preschoolers' executive functions contributed more than IQ to their pragmatic communicative skills during a semi-structured conversation with an experimenter (Blain-Brière, Bouchard, & Bigras, 2014). Specifically, children with higher working memory capacities were found to be more likely to generate contingent answers and produce clear utterances to their listeners; and children with higher inhibitory control exhibited a reduced degree of talkativeness and assertiveness in conversation.

In the current study, we investigated the role of memory in 8- and 10-year-old children's perspective-taking performance during language comprehension. We chose 8- and 10-year-old children as our subjects on two bases. First, on almost any contemporary account, children in this age range are believed to have the full range of basic theory of mind concepts. In particular, the 'Level-1' perspective-taking necessary for the director task is typically observed in children as young as 2.5 years (Flavell, Everett, Croft, & Flavell, 1981) or even in infants (Luo & Baillargeon, 2007). The second is that previous studies have shown robust age-related changes in performance on the director task across this age range (Dumontheil et al., 2010; Symeonidou et al., 2016; Wang et al., 2016).

Two key memory-related manipulations were employed in our study. Firstly, children were assigned to two different versions of the director task: a hidden perspective version and a visible perspective version. The hidden perspective version and the visible perspective version were designed to be minimally different, with the only difference being: in the hidden perspective version, the visual cues to the director's perspective were removed after an initial 5000ms-viewing-time, thus children had to infer and hold in mind what's in common ground and/or in privileged ground until the point of need; whereas in the visible perspective version, the visual cues remained available throughout each trial, thus the relevant perspective information could be inferred efficiently at the point of need. Secondly, we also systematically manipulated the relative size of common ground and privileged ground in the director task. In Experiment 2, the number of objects visible to both participants and the director was systematically manipulated from 3, 5 to 7, and the number of objects occluded from the director was held constant as 2, leaving the overall number of displayed objects varying from 5, 7 to 9. In Experiment 3, the overall number of objects was held constant as 12, and the number of objects in the common ground was systematically manipulated from 3, 5, 7 to 9, leaving the number of object occluded from to the director to vary from 9, 7, 5 to 3.

Based on these results from adults in Chapter 2, we hypothesized that children would be more prone to make egocentric errors in the hidden perspective version than in the visible perspective version of the director task. Secondly, we predicted that when the size of common ground was greater, participants would need to devote more cognitive resources to managing the increasing size of common ground, and would thus become more prone to egocentric errors and slower to respond. Thirdly, we expected 10-year-olds to make fewer egocentric errors than 8-year-olds on the experimental trials of the director task. Critically, if 10-year-olds were indeed better perspective takers than 8-year-olds, then it would be of interest to examine whether 10-year-olds and 8-year-olds were equally affected by the memory load from the hidden perspective version of the task and the size of common ground. If a factor influenced egocentrism less in older children than in younger children, then this factor would be a potential cause of developmental improvement in children's abilities to use a speaker's perspective information during communication. Specifically, by observing whether any interaction was found between age and memory-related factors (i.e., task version, common ground size) on children's degree of egocentrism, we would be able to know whether our memory-related manipulations tapped into any developmental factor that contributed to age-related improvement in perspective use during communication.

## 3.2. Experiment 2

### 3.2.1. Methods

#### *3.2.1.1. Participants*

Thirty-nine 8-year-old children (mean age 8.17, age range 7.54 to 8.71, 18 males and 21 females) and fifty-seven 10-year-old children (mean age 10.13, age range 9.53 to 10.79, 28 males and 29 females) from three primary schools located in Birmingham and neighbouring area took part in the study. Children from each class were selected according to an alphabetical class list, and were assigned to either the hidden perspective version or the visible perspective version alternately. This method ensured that condition assignment was unsystematic with respect to age, sex and any other relevant variables. Nineteen 8-year-olds took part in the hidden perspective version; another 20 took part in the visible perspective version. Twenty-nine 10-year-olds took part in the hidden perspective version, and another 28 took part in the visible perspective version. There was no difference between the ages of children assigned to the two task versions within each age group, $p = .546$ for the 8-year-old group and .751 for 10-year-old group respectively.

#### *3.2.1.2. Design and procedure*

A 2 x 2 x 2 x 3 mixed design was constructed with age (8- and 10-year-olds) and task version (hidden perspective and visible perspective version) as two between-subject factors, and condition (experimental and control conditions) and magnitude of common ground (3 objects, 5 objects and 7 objects) as two within-subject factors.

Our computerised referential communication task was introduced as an engaging computer game to children. Each child took part in the experiment individually. Children were greeted warmly, instructed step-by-step using a PowerPoint presentation and then tested individually outside of their classrooms by the experimenter. The whole experiment, consisting of an instruction phase, a practice phase and a test phase, lasted for 15-20 minutes for each individual. The experiment was presented on a 15.6' Samsung laptop. The practice and test trials were presented with Experiment Builder (SR Research Ltd, Mississauga, Ontario, Canada). Accuracy and response time data were automatically recorded by the Experiment Builder software.

**Instruction phase**

In the instruction phase, a 4 x 4 shelf image was presented on the computer screen. A number of objects were placed on the shelf. A female director, who was introduced to children as Sally, was standing behind the shelf and facing the participants. Some slots were blocked by green squared backgrounds from Sally's perspective. The experimenter explained to children that Sally could not see the objects in the blocked slots, because she was standing on the other side of the shelf and the green backgrounds were blocking her view. An image of the back of the shelf was shown to children in order to highlight Sally's limited perspective. Children were then asked five check questions about whether Sally could see certain objects. We included 3 objects from open slots and 2 objects from blocked shots into the check questions in order to make sure that children fully understood Sally's limited perspective. Children were only allowed to carry on if they answered all five questions correctly, otherwise the experimenter would again explain how Sally's perspective is limited. If children correctly answered the check questions, they were further instructed that if Sally could not see objects in the blocked slots, then she could not know about those objects, and therefore would not ask them to move any of those objects. Children's understanding of the instructions was checked by another five questions on whether Sally could know about certain objects (3 from open slots and 2 from blocked slots), and three additional questions on whether Sally could ask them to move certain objects (2 from open slots and 1 from blocked slots). Every participating child managed to answer all these questions correctly.

After checking children's understanding of Sally's limited knowledge of the shelf, one example image of the shelf (see the left image in Figure 3.1) was shown to children. Children assigned to the hidden perspective version were given the instruction that 'you have 5 seconds to remember which objects Sally does and doesn't know about. After 5 seconds, all of the slots will be blocked by a green background'; while children assigned to the visible perspective version were told that 'you have 5 seconds to look at what object Sally does and doesn't know about. After 5 seconds, Sally will start to give you some instructions.' After five seconds, children in the hidden perspective version were shown the shelf image with a covering green background (the right image in Figure 3.1) and those in the visible perspective version were shown the original image (the left image in Figure 3.1).

Each child was first given an example instruction '*nudge the ball one slot up'*, shown how to drag the ball one slot up and then drop it using the computer mouse, and then asked to practice and carry out the same 'drag and drop' action. Children were then given the next instruction '*nudge the big hat one slot down*' and were prompted to point out the two hats Sally knew about (one big and one small). If children wrongly pointed to a hat in the blocked slot, the experimenter would point to the image and say, 'look here, Sally doesn't know about this hat, because this one is in the green slot and it's blocked off'. The same instruction would then be repeated until children got it right. After identifying the two hats Sally knew about, children were asked by the question 'so which one is the big hat that Sally is talking about'. If children responded correctly, the experimenter would ask children the reason why they chose that hat, to make sure that the correct response was not driven by unexpected strategies. If children responded incorrectly, the experimenter would explicitly tell children about the way in which Sally's limited perspective constrained reference by saying 'Sally couldn't see this red hat, so she does not know that this red hat is on the shelf, and therefore she can't be talking about this red hat. Instead, Sally must be talking about this blue hat, which she can see.' At the end of the instruction phase, another shelf image with a different array of objects was presented to children as a comprehension check. A critical instruction '*nudge the big bowling pin one slot right*' was delivered. Children who not only failed the comprehension check in the instruction phase but also performed at floor level in the main test phase were excluded prior to the analysis (see section 3.2.2 for details).

Figure 3.1. Examples of shelf images used for the visible perspective version and hidden perspective version respectively.

**Test phase**

Children undertook either the hidden perspective or the visible perspective task version they were assigned to. For each task version, two practice shelf images were first presented with 3 instructions each. Twenty-four test shelf images were then presented with 2-3 instructions each. Twelve shelf images corresponded to the experimental condition, and the other 12 shelf images corresponded to the control condition. In the experimental condition, the object which best fitted a critical instruction from the participants' point of view was always in the participants' privileged ground, thus made it a 'distractor' from the target referent. Participants had to use the information about the perspective difference between the director and participants, in order to select the correct referents. Control trial images were minor adaptations of each corresponding experimental trial image, whereby the competing object in ('distractor') in the participants' privileged ground was replaced by an irrelevant object, which did not compete with the target referent as a potential referent. As illustrated in

Figure 3.2, the pink balloon in the privileged ground in the experimental condition was replaced by the broccoli when in the control condition. Therefore, in the control condition, the use of the director's perspective was not a prerequisite for correctly selecting the target referent. We checked the visual salience (e.g., size and colour) of the 'distractors' in the experimental condition versus the irrelevant objects which replaced the distractors in the corresponding control condition: the distractors in the experimental condition was not systematically more visually salient than the irrelevant object in the corresponding control condition, and vice versa.[2]

The shelf image in the control condition was presented at least 6 shelf images apart (which equals to at least 12 instructions apart) from its corresponding shelf image in the experimental condition. 24 shelf images were grouped into 4 blocks, each block was mixed of experimental and control trials. Participants could take breaks between blocks.

**Experimental condition** **Control condition**

Figure 3.2. Example images for an experimental condition trial (left) and its corresponding control condition trail (right).

---

[2] Future studies using this experimental paradigm could consider creating two sets of stimuli by balancing the visual salience of the distractor and the irrelevant object for each shelf image, e.g., by swapping the colors of the distractor and the irrelevant object.

The shelf image in the experimental condition and the shelf image in its corresponding control condition shared the same set of 2-3 verbal instructions, one of which was a critical instruction, and the others were fillers. The position of the critical instruction varied between the first and third in the sequence. Each instruction was made of individually pre-recorded words in order to prevent participants from using co-articulation to identify an object prior to the onset of the noun. The structure of the critical instructions was 'nudge the [scalar adjective] [noun] one slot [directional word]' (e.g., nudge the small ball one slot up). The structure of the fillers was either 'nudge the [scalar/normal adjective] [noun] one slot [directional word]' or 'nudge the [noun] one slot [directional word]'. Directional words included 'left' 'right' 'up' and 'down'. Children were told that the directional words 'left' and 'right' referred to their own left and right sides, and labels indicating children's left and right sides were presented across the top of the laptop screen during their participation. For critical instructions, only 'up' and 'down' directional words were used.

Children were instructed to respond as accurately and as quickly as possible. If children did not respond within 8500ms after the onset of the adjective or noun, then the trial would timeout, bringing up the next instruction or the next shelf image.

### 3.2.2. Results

We performed both a comprehension check (i.e. whether children responded correctly to a critical question at the end of instruction phase), and a floor performance check (i.e., whether children responded correctly on more than 2 trials out of 12 trials) on participating children's performance. The purpose of adopting these two criteria together for exclusion was to exclude children whose low performance was highly likely due to the lack of understanding of the current experiment instructions, rather than egocentrism. For children who failed to pass the comprehension check question in the instruction phase, it was not clear whether this failure occurred only accidently or it truly reflected insufficient understanding of the task instructions. For children who failed to perform significantly above floor, it was not clear whether their low performance was the true reflection of their insufficient understanding of the task instructions, or was induced by the cognitively demanding nature of the current experimental manipulation. Therefore, we used both rules to exclude participants. Two 8-year-olds (which were not included into the 39 eight-year-old participating children reported in section 3.2.1.1) were excluded prior the analysis due to both the failure to pass the comprehension check and the failure to perform significantly above floor.

#### *3.2.2.1. Proportion egocentric errors*

Only data from critical trials were entered into our analysis. A response was considered as correct when children selected a target object which was visible to both the director and themselves. An egocentric error referred to the selection of the distractor in privileged ground rather than the target object in the common ground. A non-egocentric error referred to the selection of other objects or spaces other than the target and the distractor. The causes of the non-egocentric errors (which constituted 5.92% of the data from the hidden perspective version in 8-year-olds, 7.92% of the data from the visible perspective version in 8-year-olds, 4.17% of the data from the hidden perspective version in 10-year-olds, and 3.72% of the data from the visible perspective version in 10-year-olds) were difficult to interpret and were not the central interest of the current study, therefore the non-egocentric errors were excluded prior to the following analysis. Trials with response timeout (1.82%) were also excluded from the analysis.

The proportion egocentric errors in the experimental condition and the control condition were calculated separately for each magnitude condition for each participant. In the experimental condition, the proportion egocentric errors of individual participants ranged from no errors to 1 in a given magnitude condition and from no errors to .917 overall. In the control condition, the mean proportion egocentric errors for each age group in each task version were all below .005. Due to the floor level of proportion egocentric errors in the control condition and the unequal variance between the experimental and control conditions, it was questionable to include the factor of condition (experimental and control) into an omnibus analysis. Therefore, a 2 x 2 x 3 mixed ANOVA was conducted for the experimental condition only, with age and task version as two between-subject factors, and common ground size as a within-subject factor. The pattern observed here in the 2 x 2 x 3 ANOVA still held for the analysis of proportion egocentric errors when we included condition (experimental, control) as a factor in a 2 x 2 x 2 x 3 mixed ANOVA.

There was a significant main effect of common ground size, $F(2, 184) = 14.026$, $p < .001$, $\eta p^2 = .132$, with a significant linear trend only, $F(1, 92) = 21.568$, $p < .001$, $\eta p^2 = .190$ (the quadratic trend was not significant, $p = .116$). This showed that the rates of egocentric errors participants made increased proportionally as the common ground size grew. A significant main effect of age was observed, $F(1, 92) = 9.314$, $p =.003$, $\eta p^2 = .092$, with 10-year-olds outperforming 8-year-olds. A significant main effect of task version was also observed, $F(1, 92) = 4.133$, $p = .045$, $\eta p^2 = .043$, with children revealing higher degrees of egocentrism in the hidden perspective version than in the visible perspective version. A significant interaction between task and age was observed, $F(1, 92) = 7.434$, $p = .008$, $\eta p^2 = .075$, and also a significant interaction between common ground size and age, $F(2, 184) = 3.627$, $p= .029$, $\eta p^2 = .038$. There was also a marginally significant interaction between common ground size and task version, $F(2, 184) = 2.740$, $p = .058$, $\eta p^2 = .031$. Most importantly, a significant 3-way interaction of common ground size x task version x age was observed, $F(2, 184) = 7.740$, $p = .001$, $\eta p^2 = .078$ (see Figure 3.3).

To follow up the significant 3-way interaction of common ground size x task version x age, two 2-way mixed ANOVAs were conducted for 8-year-olds and 10-year-olds separately, with task version and magnitude as the two factors. For 8-year-olds, there was a significant main effect of common ground size, $F(2, 74) = 10.179$, $p < .001$, $\eta p^2 = .216$, with a significant linear trend only, $F(1, 37) = 16.118$, $p < .001$, $\eta p^2 = .303$ (quadratic trend, $p = .444$). This showed that the proportion egocentric errors 8-year-olds committed increased proportionally with the increasing common ground size. A significant main effect of task version was also found, $F(1, 37) = 8.771$, $p = .005$, $\eta p^2 = .192$, as 8-year-olds made more egocentric errors in the hidden perspective version compared to the visible perspective version. Critically, a significant interaction between common ground size and task version was also found, $F(2, 74) = 6.442$, $p = .003$, $\eta p^2 = .148$. Two sets of one way repeated measures ANOVAs with common ground size as the within-subject factor were conducted for memory and visible perspective versions, respectively. For the hidden perspective version, a main effect of magnitude was observed, $F(2, 36) = 13.420$, $p < .001$, $\eta p^2 = .427$, with a significant linear trend only, $F(1, 18) = 16.753$, $p = .001$, $\eta p^2 = .482$ (quadratic trend, $p = .195$). For the visible perspective version, only a marginally significant main effect of magnitude was observed, $F(2, 38) = 2.473$, $p = .098$, $\eta p^2 = .115$.

For 10-year-olds, a 2-way mixed ANOVA was also conducted with task version and common ground size as factors. There was a significant main effect of common ground size, $F(2, 110) = 3.252$, $p = .042$, $\eta p^2 = .056$, with a marginally significant linear trend, $F(1, 55) = 3.778$, $p = .057$, $\eta p^2 = .064$ (quadratic trend, $p = .120$). No main effect of task version was observed, $F(1, 55) = .315$, $p = .577$, $\eta p^2 = .006$. No significant interaction was found between task version and common ground size, $F(2, 110) = 1.245$, $p = .292$, $\eta p^2 = .022$. [3]

---

[3] The proportion egocentric error data suffered from the problem with non-normal distribution. We considered the concerns around running ANOVA on proportion data when the data violated normal distribution. Although many studies give strong support for the robustness of the ANOVA under application of non-normally distributed data (e.g., Glass, Peckham, & Sanders, 1972; Harwell, Rubinstein, Hayes, & Olds, 1992; Schmider, Ziegler, Danay, Beyer, & Bühner, 2010), we also conducted GEE (generalised estimating equation) analysis to double check the effects observed in ANOVAs. GEE is an estimating method

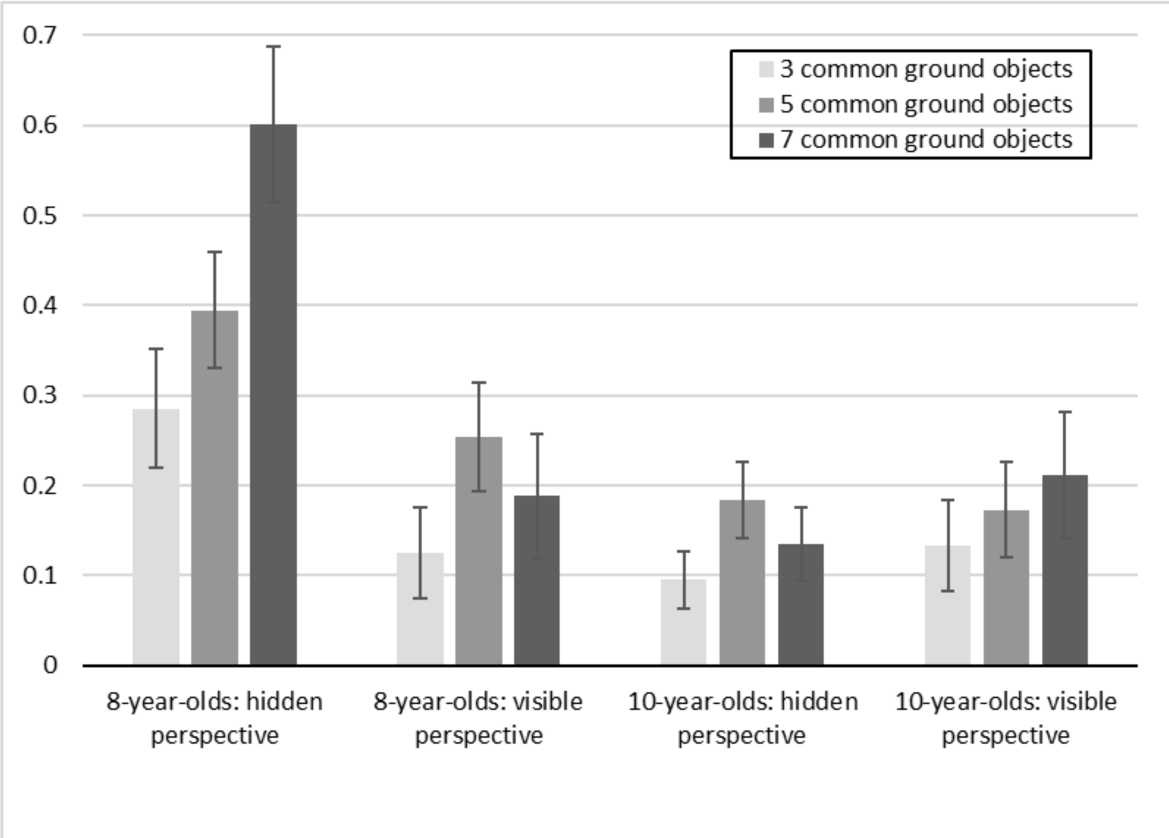


Figure 3.3. Proportions of egocentric errors on experimental trails in Experiment 2. Error bars show standard errors.

#### *3.2.2.2. Response time*

Only trials with correct responses were included in the response time analysis. As the error rates among the experimental trials were uneven among different task versions in different age groups, the numbers of data points across conditions varied. The response time data therefore could not be interpreted with confidence, as it was likely for the results to be

---

extended from the linear model used for ANOVA, which accommodates non-normal distribution, repeated measurements and categorical independent variables. This GEE approach has not been adopted in our main analysis, as it provides a less complete treatment of the data. But in the complimentary GEE analysis, GEE results converged with the results from ANOVA on proportion egocentric errors in Experiment 2. The magnitude of common ground was specified as a within-subject variable, age and task version as two between-subject variables, and percentage egocentric errors as the dependent variable. The distribution of the dependent variable was specified as poison rather than normal distribution. Convergingly, the GEE results showed a significant main effect of magnitude ($p < .001$), a main effect of age ($p = .002$), a main effect of task version ($p = .042$) and a significant three-way interaction of magnitude x age x task version ($p =. 002$).

inflated by the variance induced by insufficient data points. For this reason, we only reported the descriptive statistics for the response time (see Figure 3.4). Figure 3.4 demonstrates that overall children tended to spend longer time to make a correct response in the experimental condition than in the control condition for each magnitude condition. Children tended to spend longer time to make a correct response when there were more objects in the common ground. This effect of common ground size seemed to be prominent for both the hidden perspective task and visible perspective task, and also for both 8-year-olds and 10-year-olds, with the only exception being the control condition trials in visible perspective task for 10-year-olds. Note that these were only trends observed from the descriptive statistics but not from formal statistical analysis, therefore could not be further interpreted with confidence.

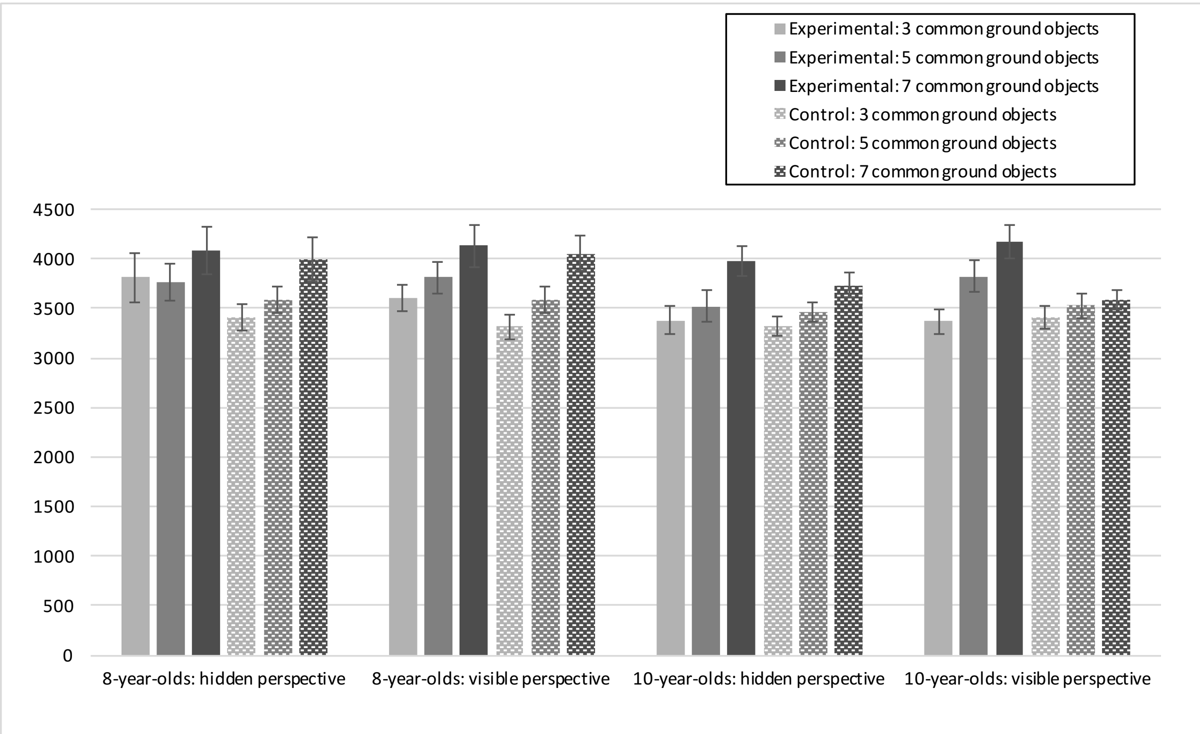


Figure 3.4. Means and standard errors of response time (ms) for 8 and 10-year-olds in Experiment 2.

### 3.2.3. Summary

Overall, children performed more egocentrically when they were under the enforced memory load in the hidden perspective version, and when the common ground size was greater. As expected, 10-year-olds committed fewer egocentric errors compared to 8-year-olds. Critically, an interaction was observed between the effects of task version, common ground size and age on egocentrism. In 8-year-olds, we found a systematic linear effect of common ground size on egocentrism in the hidden perspective version but not in the visible perspective version; while for 10-year-olds, there was limited evidence suggesting that common ground size had a systematic linear effect on their degree of egocentrism, and no difference was found between the hidden perspective and visible perspective versions. This interaction with age suggests that younger children were disproportionally affected by the common ground size when they were under the enforced memory load than when such memory load was absent. The results pattern from Figure 3.3 is also consistent in suggesting that the main effect of task version, age and the size of common ground were mainly driven by the distinctive performance of 8-year-old children in the hidden perspective version. Altogether, this suggests that our memory manipulations might have partially accounted for the age-related improvements between 8- and 10-year-olds' perspective-taking performance.

However, two confounding variables were associated with the experimental design in Experiment 2. The first was that although we were manipulating the number of objects in the common ground, the number of objects in the privileged ground was held constant as 2, which left the total number of objects increasing with the size of common ground. The greater the common ground size, the greater the whole object array size, and the more complex the object array. Frick et al. (2014) showed that the visual complexity of an array of objects influenced 5- to 8-year-old children's visuo-spatial perspective-taking performance. Therefore, it was unclear from Experiment 2 whether 8-year-old children's degrees of egocentrism in the hidden perspective version varied as a function of the size of common ground as we intended to examine, or alternatively as a function of the overall array size or visual complexity. Secondly, the size of privileged ground was held constant as 2 in quantity. If children in the hidden perspective version were being strategic, then they could choose to only encode the two objects in the privileged ground regardless of the number of objects presented in the common ground. Therefore, the cognitive load of encoding and holding in mind relevant perspective information could have been greatly reduced and kept invariable across different magnitude conditions. Given that only 8-year-old children in the hidden perspective version were affected by the increasing common ground size, it is plausible that the 8-year-olds attempted to encode the objects in the common ground or the whole object array, whereas 10-year-olds may have identified this efficient encoding strategy and encoded the objects in the privileged ground instead.

In Experiment 3, we aimed to disentangle the effect of common ground size from the confounding effect of object array size and also to rule out the use of potential memory strategy. We systematically varied the number of objects in the common ground between 3, 5, 7 and 9 whilst holding the total number of objects constant as 12, leaving the number of objects in the privileged ground to vary between 9, 7, 5 and 3.

## 3.3. Experiment 3

### 3.3.1. Methods

#### *3.3.1.1. Participants*

Sixty-two 8-year-old children (mean age 7.91, age range 7.17 to 8.81, 30 males and 32 females) and 71 10-year-old children (mean age 9.97, age range 9.18 to 10.81, 36 males and 35 females) from three primary schools located in the Birmingham area took part in the study. Children from each class were selected according to an alphabetical class list, and were assigned to either the hidden perspective version or the visible perspective version alternately. Twenty-nine 8-year-olds took part in the hidden perspective version; another 33 took part in the visible perspective version. Thirty-four 10-year-olds took part in the hidden perspective version, and another 37 took part in the visible perspective version. There was no difference between the ages of children assigned to the hidden perspective version and those assigned to the visible perspective version within each age group, p =.380 for 8-year-olds, and p = .725 for 10-year-olds. Six additional 8-year-olds and 5 additional 10-year-olds were excluded prior the analysis due to both their failure to pass a comprehension check at the end of the instruction phase and also failure to perform significantly above floor level.

#### *3.3.1.2. Design and procedure*

A 2 x 2 x 2 x 4 mixed design was constructed with age (8- and 10-year-olds) and task version (hidden perspective and visible perspective versions) as between-subject factors, and condition (experimental, control) and the magnitude of common ground (3, 5, 7, and 9 common ground objects) as within-subject factors.

The design and procedure were identical to that of Experiment 2 with the following exceptions. All shelf images contained 12 objects in total, while the number of objects in the common ground varied between 3, 5, 7, and 9, thus leaving the number of objects in the privileged ground to vary between 9, 7, 5, and 3. A total of 32 shelf images were presented. Sixteen shelf images were used for the experimental condition with 4 shelf images for each magnitude condition, and the other 16 shelf images were for the control condition also with 4 for each magnitude condition.[4] Each shelf image was accompanied by 2 verbal instructions,

[4] In the hidden perspective version, a technical error occurred in one of the control trials, whereby an image corresponding to its matching experimental condition was presented in the 5000ms viewing time before a correct image onset and remained onscreen during the verbal instructions. Therefore, the data from this control trial in both the hidden perspective and

one of which was a critical instruction, and the other instruction was a filler instruction. The critical instruction could occur as either the first or the second instruction in a sequence. Thirty-two shelf images were grouped into 4 blocks, and each block had 8 shelf images, mixed of experimental and control trials. The shelf image in the control condition was presented at least 8 shelf images apart from its corresponding shelf image in the experimental condition. One practice block consisting of 3 shelf images was presented prior to the test phase, each shelf image was accompanied by 2 verbal instructions.

## 3.3.2. Results

### *3.3.2.1. Proportion of egocentric errors*

Non-egocentric errors (which constituted 10.46% of the data from the hidden perspective version in 8-year-olds, 11.14% of the data from the visible perspective version in 8-year-olds, 9.11% of the data from the hidden perspective version in 10-year-olds, and 7.85% of the data from the visible perspective version in 10-year-olds) and response timeouts (2.79% overall) were excluded prior to the analysis. In the experimental condition, the proportion egocentric errors of individual participants ranged from no errors to 1 in a given magnitude condition and from no errors to .854 overall. In the control condition, the mean proportion egocentric errors for each age group in each magnitude condition of each task version were all below .011.

The data from the experimental condition were submitted to a 2 x 2 x 4 mixed ANOVA with age and task version as two between-subject factors, and magnitude of common ground (3, 5, 7, 9 common ground objects; referred to as 'magnitude') as a within-subject factor. There was a significant main effect of the magnitude of common ground, $F(3, 387) = 14.980$, $p < .001$, $\eta p^2 = .104$, with a significant linear trend only, $F(1, 129) = 38.001$, $p < .001$, $\eta p^2 =$

---

visible perspective versions were excluded prior to the analysis, in order to make fair comparisons between children's performance in these two task versions.

.228 (quadratic trend, $p$ = .374, and cubic trends, $p$ = .227). This suggested that with the increase of the size of privileged ground (and correspondingly the decrease of the size of common ground), accuracy increased in a systematic linear way. A significant main effect of task version was also observed, $F(1, 129) = 24.587$, $p < .001$, $\eta p^2 = .160$, with participants in hidden perspective version committing more egocentric errors than those in the visible perspective version. A marginally significant main effect of age was also observed, $F(1, 129) = 3.489$, $p = .064$, $\eta p^2 = .026$. In general, 10-year-olds showed a tendency to outperform 8-year-olds, but this tendency did not reach statistical significance. A marginally significant interaction between age and task version was also observed, $F(1, 129) = 3.194$, $p = .076$, $\eta p^2 = .024$. No other significant 2-way interaction was found, $p$s > .100. The 3-way interaction was also non-significant, $F(3, 387) = 1.699$, $p = .167$, $\eta p^2 = .013$. The results can be seen in Figure 3.5.[5]

[5] GEE (generalised estimating equation) results converged with the results from ANOVA on proportion egocentric errors in Experiment 2. In the complimentary GEE analysis, the magnitude of common ground was specified as a within-subject variable, age and task version as two between-subject variables, and percentage egocentric errors as the dependent variable. The distribution of the dependent variable was specified as poison rather than normal distribution. Convergingly, the GEE results showed a significant main effect of magnitude ($p$ < .001), a marginally significant effect of age ($p$ = .053), a main effect of task version ($p$ < .001) and a marginally significant interaction of age x task version ( $p$ = .064).

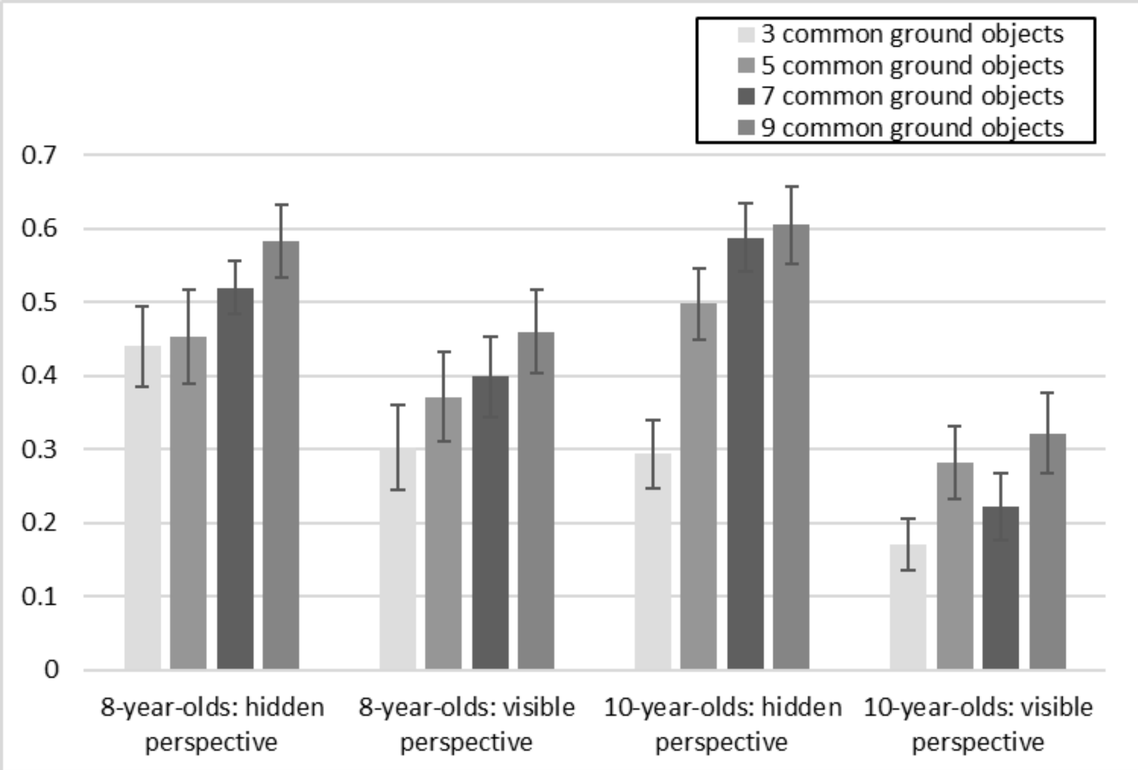


Figure 3.5. Proportions of egocentric errors on experimental trails in Experiment 3. Error bars show standard errors.

Because of its potential importance, we followed up the interaction between age and task version, even though it only approached statistical significance. Two independent t-tests were first conducted between the hidden and visible perspective versions for 8- and 10-year-olds separately. Eight-year-olds made significantly more egocentric errors in the hidden perspective version than in the visible perspective version, $t(60) = 2.216$, $p = .030$, *Cohen's d* =.571. Similarly, 10-year-olds also performed more egocentrically in the hidden perspective version than in the visible perspective version, $t(69) = 5.013$, $p < .001$, *Cohen's d* = 1.192. This tendency confirmed the main effect of task version on egocentrism. Two independent t-tests were then conducted between 8- and 10-year-olds for the hidden perspective version and the visible perspective version separately. Two age group performed similarly for the hidden perspective version, $t(61) = .033$, $p = .974$, *Cohen's d* = .009; but for the visible perspective version, 10-year-olds were significantly less egocentric compared to 8-year-olds, $t(68)$ =

2.530, $p$ = .014, *Cohen's d* = .604. It seems that the marginally significant effect of age and the marginally significant interaction between effects of age and task version was mostly driven by the difference between 8-year-olds' performance and 10-year-olds' performance in the visible perspective task version[6].

### *3.3.2.2. Response time*

As for Experiment 2, only the descriptive statistics of the response time for Experiment 3 were reported in Figure 3.6. No clear and systematic trends could be observed from the descriptive statistics in Figure 3.6.

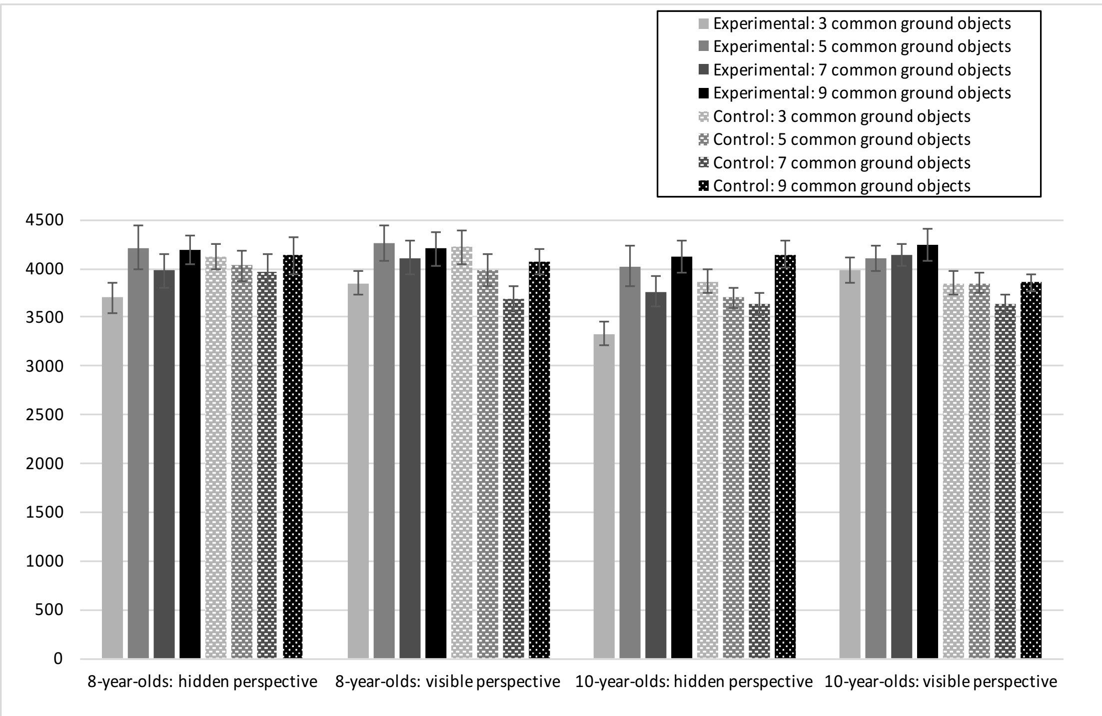


Figure 3.6. Means and standard errors of response time (ms) for 8 and 10-year-olds in Experiment 3.

---

6 Experiment 3 was first run with just 92 of these children. Analysis of data from this sample yielded the same results pattern to the results we presented in this paper, i.e., significant main effects of task version and common ground size, marginally significant effect of age, and marginally significant interaction between age and common ground. To check whether the marginally significant effects in our initial sample were due to insufficient power, we extended the sample size by 50% to gain the full sample reported here, but this did not change the pattern of results.

### 3.3.3. Summary

Experiment 3 provided convergent evidence with Experiment 2 on the significant roles of the number of items in common ground and the need to retain these objects in memory while interpreting the director's instructions. When the total number of objects was kept constant, the size of common ground still systematically affected children's perspective-taking performance. The greater the common ground size, the more egocentric errors children committed. Moreover, we found that children were more prone to select the distractor when they were under the enforced memory load than when such memory load was not enforced. Ten-year-old children showed a tendency to be more successful at identifying the target referent in the common ground compared to 8-year-olds, but the difference was only marginally significant. Moreover, we observed a marginally significant interaction between age and task version, which was mostly driven by 10-year-olds performing less egocentrically than 8-year-olds in the visible perspective version, but equally egocentrically in the hidden perspective version. The implications of these findings were discussed in details in the next section.

## 3.4. General discussion

The aim of the present study was to examine whether memory factors could explain the variations in children's perspective-taking performance. Eight- and 10-year-old children were instructed to take a speaker's perspective either under enforced memory load (i.e., the hidden perspective version) or under reduced memory load (i.e., the visible perspective version). In Experiment 2, the size of common ground was systematically varied with the size of privileged ground held constant as 2; in Experiment 3, the size of common ground was systematically varied with the total number of objects held constant as 12. Consistent with previous studies, egocentric effects were clearly observed in the measure of errors in our study (e.g., Apperly et al., 2010; Dumontheil et al., 2010; Keysar et al., 2003; Nilsen & Graham, 2009). The current study had three key objectives: (a) to evaluate the effect of holding in mind a speaker's perspective on children's degree of egocentrism, (b) to assess the effect of common ground size on children's degree of egocentrism, (c) to examine whether 10-year-old children would perform less egocentrically compared to 8-year-olds in our experimental setting, and whether the manipulated memory factors would drive this age-related reduction in egocentrism. The current findings will be discussed in further details below in relation to these three objectives.

Our study was the first attempt to directly measure the cost of holding in mind a speaker's perspective on participants' performance in the director task by removing the visual cues to the director's perspective after a brief viewing time. Our findings from both experiments support the conclusion that children were more egocentric in the hidden perspective version than in the visible perspective version. The specific cognitive demands of each task can be decomposed as follows (Apperly et al., 2010). In the hidden perspective version, participants had to encode, retain and later integrate the director's perspective with her instructions in order to successfully resolve the reference. Failure to include the target referent into the memory record of potential common ground referents would lead to failure to correctly recognise the target referent. However, in the visible perspective version, participants could identify the director's perspective and integrate it with her verbal message at the point of need but it was not necessary to store relevant perspective information for any longer than it took to infer and integrate perspective information. The only difference between the two task versions was that there was an extra memory load of remembering and retaining encoded perspective information in the hidden perspective version. Therefore, it seems reasonable to conclude that the increase in egocentric errors arose from the difficulty children face in managing this extra memory load in the hidden perspective version. This finding corresponds to the everyday situations in which communication requires us to *remember* who knows what, rather than figure out who knows what at the precise time we need this information. Our study operationalises a way to incorporate this memory demand which is frequently and naturally present in our everyday communication into the widely-used director task, and results confirm that this memory load could partially explains children's variations in perspective use in different real-life situations.

Secondly, our study systematically manipulated the size of common ground in two different ways: in Experiment 2, we manipulated the number of objects in the common ground between 3, 5 and 7 whilst keeping the number of objects in the privileged ground constant as 2; In Experiment 3, we manipulated the number of objects in the common ground between 3, 5, 7 and 9 whilst keeping the overall number of objects as 12. Two experiments consistently found that both 8- and 10-year-olds were less successful at accommodating a speakers' limited perspective in communication when there was a greater amount of common ground information. This effect of increasing common ground size on egocentrism is in line with a previous study showing the same effect of common ground size on adults' egocentrism under the enforced memory load of holding in mind a speaker's perspective (Zhao, Wang, & Apperly, under review). While common ground size co-varied with the total number of objects in Experiment 2, this was not the case in Experiment 3. Frick et al.'s study (2014) suggested that visual complexity of the layouts of object array had a detrimental effect on 5- to 8-year-olds' perspective-taking performance when their own perspective differed from another person. Our findings have advanced knowledge in this field by providing evidence that even when two object arrays were of the same degree of visual complexity, the size of common ground still affected children's degree of egocentrism. These findings provide support for the long assumed notion that listeners work on the basis of common ground to interpret a speaker's message (Clark & Marshall, 1978, 1981), and have further extended this assumption from adult communicators to young communicators who are generally less experienced in communication and are equipped with limited social and cognitive capacities. Our findings are consistent with Nadig and Sedivy (2002)'s findings that children do show considerable sensitivity to common ground information in referential communication, although our study is not designed to examine the time course regarding this sensitivity to common ground. The findings also accord with Barr (2008)'s claim that people do anticipate items in common ground to be referred to, but also suffer from egocentric interference when they need to integrate linguistic input with perspective information. More importantly, our findings provide evidence about how the cognitive demand of managing common ground size may constrain people's ability to use other's perspective to guide communication.

Thirdly, we sought evidence for age-related improvement in children's ability to adopt a speaker's perspective in referential communication, and also sought the possible cognitive factors that drove this age-related improvement. Results from Experiment 2 demonstrated that 10-year-olds outperformed 8-year-olds in the director task in the proportion of egocentric errors they committed. This finding is consistent with a number of previous studies suggesting that children's ability to use other people's mental states continues to grow over time after they have acquired the basic theory of mind competence (Dumontheil et al., 2010; Epley, Morewedge, et al., 2004; Symeonidou et al., 2016), especially in accord with Wang et al. (2016) showing that 10-year-olds committed significantly fewer egocentric errors than 8-year-olds in the director task. Experiment 2 also witnessed an interaction among factors of task version, common ground size and age. The performance of 8-year-olds in the hidden perspective version was disproportionally affected by the increasing common ground size compared to 8-year-olds in the visible perspective version and 10-year-olds in both versions of the task. This provides some evidence that the memory-related factors we manipulated in

Experiment 2 might have contributed to the age-related improvement between 8- and 10-year-olds' perspective-taking performances. However, as discussed in section 3.2.3, this significant effect of age and significant interaction of age x task version x common ground size on egocentrism in Experiment 2 might have been driven by older children's improved ability of identifying and employing the appropriate memory strategies when they were under the extra memory load of remembering the director's perspective, rather than older children's improved memory capacity of correctly remembering a larger amount of perspective information. This speculation is consistent with the suggestion that one key mechanism of the development of visuo-spatial working memory is that older children are better at recognising and adopting possible processing strategies (see Pickering, 2001 for a review paper). After controlling for the possibility of being strategic in only encoding the objects in privileged ground, Experiment 3 found that 10-year-olds committed only marginally fewer egocentric errors compared with 8-year-olds. Moreover, only a marginally significant interaction was found between the factors of age and task version in Experiment 3. Specifically, both age group showed a similar degree of egocentrism in the hidden perspective version, but 10-year-olds showed a tendency to perform less egocentrically than 8-year-olds in the visible perspective version.

Overall, the findings from Experiment 2 and Experiment 3 together provided some evidence for an age-related development in children's ability to accommodate a speaker's limited perspective during referential communication. While 10-year-olds performed significantly less egocentrically than 8-year-olds in Experiment 2, this age-related effect was marginally significant in Experiment 3. As our study only tested 8- and 10-year-old children, it is possible that this limited age range may have obscured the age-related development we intended to find. It would be sensible for future research to have a larger age-range when it comes to examining age-related development in perspective use.

Critically, the current study examined whether either of the two memory-related factors (i.e., remembering the director's perspective and managing common ground size) would exert less effect on older children's degree of egocentrism compared to younger children's. The pattern observed in Experiment 2 was consistent with 10-year-olds being better than 8-year-olds at managing the memory demand of remembering the increasing common ground size in the hidden perspective version but not in the visible perspective version. However, after eliminating the possibility of employing a useful encoding strategy, Experiment 3 only found a marginally significant interaction between task version and age, whereby 10-year-olds were less egocentric than 8-year-olds in the visible perspective version but not in the hidden perspective version. One plausible explanation was that 10-year-olds were more able than 8-year-olds to infer the director's perspective and use it to guide comprehension when the need to remember the director's perspective was absent in the visible perspective version, but these 10-year-olds were no better than 8-year-olds at reliably remembering the director's perspective in the hidden perspective version. Following this interpretation, however, we would expect the same pattern of results to have also emerged between the two age groups from the visible perspective version in Experiment 2, which was not the case. It is unclear from the present study why interactions involving the factor of task version showed different patterns in Experiment 2 and Experiment 3. Therefore, this proposed explanation should be interpreted with caution and may need further exploration.

Importantly, although common ground size had a systematic effect on children's degree of egocentrism in both Experiment 2 and Experiment 3, no interaction involving common ground size was found in Experiment 3 after controlling for the aforementioned confounding variable. This finding may appear to be surprising, as previous studies have found that children's successful performance on theory of mind tasks draws upon their executive functions including working memory and inhibitory control (e.g., Carlson, Moses, & Breton, 2002; Carlson, Moses, & Claxton, 2004; Davis & Pratt, 1995; Frye, Zelazo, & Palfai, 1995; Hansen Lagattuta, Sayfan, & Harvey, 2014; Lagattuta, Sayfan, & Blattman, 2010). Therefore, it seems plausible to assume that the increasing size of common ground might influence younger children's perspective-taking performance more strongly by placing an additional memory demand on their rather limited memory capacities compared to older children. However, our results provide no clear evidence for this suggestion. This finding, though surprising, is consistent with Wang et al. (2016) where they found that 8- and 10-year-olds were equally influenced by the complexity of the instructions they had to integrate with the director's limited perspective. These findings suggest that we cannot take it for granted that the cognitive factors that affect children's perspective-taking performance  are also those that drive children's age-related improvement in perspective use. In additional to the memory capacity to manage increasing common ground size examined in the current study and the capacity to manage language complexity examined in Wang et al. (2016), inhibitory control might be another potential cognitive factor that drive children's age-related improvements in perspective taking. This speculation is supported by the evidence that children's egocentric error rates on the classic director task correlated with their inhibitory control capacities (Nilsen & Graham 2009). Therefore, it is possible that older children are less egocentric than younger children because of their age-related improvements in inhibitory control capacities. Thus, age-related improvements in inhibitory control are worth exploring in future research as a source of age-related reduction in egocentrism.

To conclude, the present study incorporated into the commonly-used director task two types of memory demands which are frequently present in our everyday communication: the memory demand to hold in mind a speaker's perspective, and the amount of common ground information to keep track of. We examined the effects of these two memory demands on 8- and 10-year-old children's perspective-taking performance during language comprehension.

Findings demonstrated that 8- and 10-year-olds' rates of egocentrism were modulated both by the demand of holding in mind a speaker's perspective, and by the varied amount of common ground information they needed to keep track of. While 10-year-olds were significantly less egocentric than 8-year-olds in Experiment 2, this effect was marginal in Experiment 3. No clear evidence was found for the memory-related factors to be accountable for age-related improvement in children's perspective-taking performance. Nonetheless, the current findings imply that children will be most successful at accommodating a speaker's limited perspective when they can easily access information about what is shared between them and the speaker rather than having to remember and later retrieve this information from their memory. The findings also suggest that an increase in the number of potential referents shared between listeners and speakers may cause children to be more egocentric when they need to use a speaker's limited perspective to identify a correct referent. When talking to children, highlighting what is shared between conversational partners or any perspective difference to narrow down the scope of potential referents would help them to better avoid falling back to using their own perspective to interpret what other people are referring to.

CHAPTER IV

# Introduction to a low-level memory-based mechanism for common ground

## 4.1. Links from Chapter 2 and 3 to Chapter 5 and 6

Chapters 2 and 3 examined the role of memory in adults and children's perspective taking performance using a newly adapted referential communication task (i.e., the director task). The study found that the cost of holding in mind a speaker's perspective for a longer duration had a detrimental effect on both adults' and children's perspective taking performance. Adults with lower working memory capacities made more egocentric errors. Both adults and children made significantly more egocentric errors when the size of common ground was greater.

Findings from Chapters 2 and 3 together suggest that the abilities to form common ground representations and to make relevant theory of mind inferences during referential communication was constrained by the memory demands of the specific communication task as well as by people's working memory capacities. Memory processes are involved in the processes of perspective taking during communication: people form perspective representations about what the conversational partner knows or does not know about, store these perspective representations in their memory, and then retrieve from memory and integrate with linguistic input when these perspective representations are needed to disambiguate linguistic reference. Failure to form, store or use these perspective representations will result in committing egocentric errors.

These findings naturally lead to a series of questions. How do people form and store common ground representations in memory? How does the memory system support the efficient use of common ground representations during everyday communication? What is the nature of the common ground representations stored in memory—do they include rich social contextual information, or do they only include simple binary information which indicates whether a piece of information is known to someone or not? More critically, is it necessary for one to rely on the explicit theory of mind inferences to ensure successful use of common ground? Alternatively, can the use of common ground be achieved through submentalising, without necessarily going though explicit mental states inferences? Research in Chapter 5 and 6 was conducted to cast light on these questions.

In Chapters 2 and 3, the director task required participants to make explicit mindreading inference about the director's perspective and use relevant perspective information to identify the target referent in the common ground. In contrast, the experiences in Chapters 5 and 6 tested whether the common ground representations can be established through a low-level memory-based mechanism, which supports the encoding, storage and retrieval of common ground representations.

## 4.2. Different accounts of common ground representations

Common ground representations shape the contents of a conversation and the manner in which they are delivered. Different accounts have been proposed by researchers regarding the nature of common ground representations people form. I will review three theoretical positions on this topic.

**Rich, episodic and 'diary-like' common ground representations**

Clark and Marshall (1978, 1981) proposed that people use a *Reference Diary*, i.e., an updated record of partner-specific knowledge, to help them keep track of common ground. According to Clark and Marshall, common ground information is stored in the form of rich, episodic, diary-like representations. This means that people need to keep an elaborated and up-to-date model of their conversational partner. Moreover, people need to explicitly refer to this model on a moment-by-moment basis during communication, which may place higher processing demands than people could plausibly satisfy in daily conversation.

**'One-bit' common ground representations**

Galati and Brennan (2006) proposed a 'one-bit model' of common ground (see also Brennan & Hanna, 2009; Shintel & Keysar, 2009; and Galati & Brennan, 2010). The 'one-bit model' suggests that under some circumstances individuals can rely on a single cue (e.g., whether the director can/cannot see this object, whether something was/was not talked about) to work out whether a certain piece of information is shared with their conversational partner. This 'minimal' account of common ground can explain why in some cases people can get access to common ground representations readily and rapidly, and use these common ground representations to adapt their language processing behaviours immediately.

**An ordinary, low-level memory mechanism for common ground**

Independent from both the full-blown 'diary-like' and the minimal 'one-bit' representations of common ground, Horton and Gerrig (2005a, 2005b, 2016) proposed a relatively low-level memory mechanism for common ground representations. This memory-based approach emphasises the role of domain-general processes of memory encoding and retrieval in the representations of common ground. Specifically, they proposed that common ground can be represented via ordinary memory processes, which can be further divided into two processes: a strategic process and an automatic process. The strategic process refers to the process that interlocutors strategically/explicitly consider whether a specific piece of information is shared between each other. The automatic process is a cue-based, automatic, fast, and effort-free memory mechanism. Under this automatic memory process, speakers and listeners depend on salient memory cues in their working memory to activate associated episodic memory traces, enabling them to automatically retrieve associated common ground representations from memory. This memory-based mechanism for common ground pointed out that many instances of partner-specific effects, which may seem to occur on the basis of people's dedicated explicit inference about common ground, can be explained by a relatively spontaneous, implicit and effortless domain-general memory process (Horton & Gerrig, 2016).

## 4.3. The low-level memory mechanism for common ground representations

Much of the evidence in support of a low-level memory mechanism for common ground representations (Horton & Gerrig, 2005a, 2005b) comes from research examining the effects of *lexical entrainment* and *partner-specificity.*

When people engage in referential communication, they often repeatedly use the same term to refer to the same object. This phenomenon is termed as '*lexical entrainment'* (Brennan & Clark, 1996; Garrod & Anderson, 1987). In the previous example, 'Mr. Apple' is a lexically entrained referring term which is agreed and shared by both conversational parties. Brennan and Clark (1996) proposed that an entrained referring expression may represent *a partner-specific 'conceptual pact'*. During the process of grounding a reference, speakers and addressees are believed to develop a 'conceptual pact' which is 'a temporary agreement about how the referent is to be conceptualized' (Brennan & Clark, 1996, p. 1484). This conceptual pact is believed to be partner-specific. When the conversational partner changes, a new conceptual pact will be established and may differ from the previous one. Experiment 3 of Brennan and Clark (1996) found that after having established a 'conceptual pact' with an addressee, the speaker would stick to the original term (e.g., '*pennyloafer*') when taking to the original addressee, even when the original term was over-informative; however, when facing a new addressee, the speaker tended to switch to a simpler term (e.g., '*shoe*'). Matthews, Lieven, and Tomasello (2010) also found that even 3- to 5-year-old children were slower to react to a new term compared to an original term which was previously established in a warm-up game. More importantly, this delay was larger when participants were with the original conversational partner who had used the original term in the warm up game, compared to with a new partner. This study suggests that even 3-year-old children were sensitive to conceptual pacts during referential communication.

Recent years have seen an increase in the number of studies examining and/or supporting a low-level memory mechanism for common ground within the framework of referential communication (Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013; Horton, 2007; Horton & Gerrig, 2005b; Horton & Slaten, 2012; Matthews et al., 2010; Rubin, Brown-Schmidt, Duff, Tranel, & Cohen, 2011; Shintel & Keysar, 2007, 2009). This mechanism explains partner-specific effects in terms of partner-specific memory associations. In this account, during interaction, information about conversation partner might become associated with a wide range of related information in episodic memory (e.g., the referring term for a specific referent). Partner-based information serves as memory cues, and its presence makes the memory traces strongly associated with the partner highly accessible.

Horton and colleagues suggests that partner-specific information is retrieved under a memory mechanism named '*resonance*' (Horton, 2007; Horton & Gerrig, 2005a, 2005b). 'Resonance' was first presented by Ratcliff (1978) for explaining the process of memory retrieval. It was later regarded by Graesser et al. (1997) as an important cognitive component in discourse comprehension. Resonance is considered to be an efficient, automatic and effortless mechanism. When the contents in working memory match closely with traces in long-term memory, resonance is likely to occur between the memory cues active in working memory and the information stored in long-term memory, thus causing the information in long-term memory to be activated and retrieved.

One piece of evidence for this memory-based account came from Horton's study (2007). In this study participants performed a picture-naming task. Participants were first given the opportunity to associate specific words or object categories with each of the two experimental partners. Participants then named a series of pictures in the presence of each experimental partner. Results showed that participants were faster to name an object when faced with the same partner as in the established partner-object episodic association than when faced with a different partner. Participants later took a partner-identification test, where they were asked to indicate which partner they had seen certain words with. No correlation was found between participants' implicit partner-specific priming effect and their explicit recall of partner-object association. The lack of correlation suggested that the observed partner-specific effect did not result from explicit mindreading inferences, but from relatively implicit memory process. This study demonstrated that the associated partner identity served as a memory cue that facilitated the retrieval of the name of the object, and that the partner-specific effect occurred even in a non-communicative context where no explicit communicative goal existed.

A recent study by Gorman et al. (2013) also provided support to the low-level memory-based hypothesis of conversational common ground. The study found that shared learning experience with a partner in the study phase provided more contextual cues for participants to better distinguish between what was shared and what was privileged in a following referential communication task. Such enhancing effect of shared learning experience was even observed when the study partner was not the communication task partner. The results indicated that participants were able keep track of shared information based on memory cues, and that they would even generalise the shared information to another partner when the memory cues are strong.

Similarly, the lack of a partner-specific effect in some circumstances can be explained within the memory-based account of common ground. For example, Shintel and Keysar's (2007) study showed that participants (as listeners) expected speakers to be consistent in using the same referring expression for a referent, even when the referring expression was used previously with another listener rather than themselves. This consistency expectation was unlikely to be motivated by the assumption of cooperativeness underlying common ground inferences. Shintel and Keysar (2007) argued that when listeners formed memory representations of mappings between a referring expression and an object, they incorporated the identities of communication partners as contextual cues into their memory. When the partner-based cue is salient, 'it *can result in speaker-specific activation of existing expression–referent mappings*'. However, if the partner-based cue is not salient enough, '*it may be swamped by a relatively much stronger cue provided by the repetition of the same expression, which may explain why a speaker-specific facilitation benefit was not found*' (Shintel & Keysar, 2007, p. 368). Apart from the effect of the salience of memory cues, Horton and Gerrig (2005a) also suggested that memory errors might cause incorrect representations of common ground (e.g., when one piece of information is in the common ground with person A but is wrongly assumed to be in common ground with person B), which further lead to perspective-taking failures or miscommunication.

The studies in Chapter 5 and Chapter 6 explored such a low-level memory mechanism for common ground representations. I took a memory-based approach to test whether communicative contextual information would influence participants' memory for objects in a relatively automatic way. Two types of contextual information regarding the representations of conversational partners was incorporated into the newly-designed experiments: Chapter 5 examined the effect of the change of conversational partners on object recognition, and Chapter 6 examined the effect of the conversational partner's perspective status (i.e., common / privileged ground) on object recognition.

## 4.4. Influence of conversational partner change on memory: inspired by work on event cognition

'Events are what happens to us, what we do, what we anticipate with pleasure or dread, and what we remember with fondness or regret. Much of our behavior is guided by our understanding of events' (Radvansky & Zacks, 2014, p. 1). As interpreted by Brown-Schmidt and Duff (2016), Clark and Marshall (1981) hold an event-based view of how people gain explicit access to common ground representations. An example from Clark and Marshall (1981):

*When Ann uses the reference the man in the red shirt, Bob must find in memory an individual who fits that description . . .he must seek out an event that he can use along with certain auxiliary assumptions as the basis G for inductively inferring mutual knowledge of the identity of that man. (p. 53)*

Based on this event-based view, conversational partners search for an event that provides the necessary basis ('G') to determine whether something is in common ground or not. In this section, I will review the literature on event cognition and the memory processes involved in the representations of event models, and explain how the mechanism for event model representations could support a low-level memory mechanism for common ground representations.

### 4.4.1. Event model, event-indexing model, and event segmentation theory

***Situation model / Event model.*** The concept of 'situation model', which refers to the mental representation of the situation described by the discourse, was proposed as a relatively high level of representation during discourse comprehension (Van Dijk & Kintsch, 1983). The term 'event model' was later proposed as an extended construct of 'situation model' in a broader research area of event cognition, including text comprehension, film comprehension and different types of real-world experiences (Radvansky & Zacks, 2011). An event model is defined as the mental representation of the event which is currently being experienced or has previously been experienced.

***Event-indexing model.*** One key feature of event model is its multidimensionality. According to the *event-indexing model* (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995), people construct their mental representation of an event in long-term memory on the basis of five dimensions, namely temporality, spatiality, causality, intentionality (i.e. goals) and the protagonist. Discontinuity in any of these five dimensions could potentially introduce an event boundary to people's mentally constructed event model, causing the model to be updated to include a new event.

***Event segmentation theory.*** Based on the event indexing model, Zacks and his colleagues further proposed the event segmentation theory, which suggests that perceptual systems spontaneously and automatically segment a stream of ongoing activity into meaningful events based on the aforementioned five dimensions (e.g., Kurby & Zacks, 2008; Swallow, Zacks, & Abrams, 2009; Zacks, Speer, & Reynolds, 2009; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). People construct a mental model to represent the event being experienced and use the current stable model to anticipate an upcoming stimulus. If the features of the current event change, prediction of upcoming information will become difficult and may cause a large number of errors to occur. This transient increase of errors is detected by the perceptual systems, which send signals to update the mental representation of the current event based on the new features of input. This theory explains how our perceptual systems parse the continuous flow of experience into meaningful chunks and how we make sense of the dynamic and complex world in an efficient and effective way.

### 4.4.2. The effect of event boundary on memory encoding and retrieval

Event segmentation is considered to form the basis of cognitive control and memory encoding and retrieval. Radvansky (2012) indicated in his recent review that, 'how we break up streams of action into events influences how we think about things and what we remember later. These memories then help guide us through flow of experience, allowing us to develop expectations of what will happen and what to do next' (p.269). The way we experience and construct events shapes our memory and cognition, while in turn our memory and cognitive mechanism help us to go through continuous activity in an effective and efficient way.

Several studies on text comprehension (e.g., Glenberg, Meyer, & Lindem, 1987; Morrow, Bower, & Greenspan, 1989; Speer & Zacks, 2005) provided converging evidence that people's memory for recently mentioned objects was impaired after reading sentences that involved an event shift. For example in Speer and Zacks (2005), participants first read a sentence e.g. 'she could hear water running, and figured there must be a creek nearby', and the 'creek' in this sentence was the object which participant would later be asked to recognise. Participants then read either a following sentence in which a temporal shift was involved (e.g., 'An hour later, she was collecting wood for a fire.') or a sentence in which a temporal shift was not involved (e.g., 'A moment later, she was collecting wood for a fire'). Results showed that participants spent a longer time to read sentences involving such temporal shifts compared to the ones without temporal shift, and their memory for the previous mentioned objects was worse after reading sentences involving such temporal shifts. These findings were believed to suggest that the objects mentioned before an event boundary are removed from working memory and then enter the long-term memory. The information in the long-term memory either decays over time, or become less accessible and takes longer to retrieve. Either of these two consequences could account for participants' impaired memory for the objects preceding the temporal shifts.

Studies on comprehension of picture stories and films also support the event segmentation theory. For example, Gernsbacher (1985) used picture story to test participants' memory for pictures across events. Participants were asked to view a picture story, during which they were probed to identify and distinguish the original picture from a left-right reversed form of the original picture. Participants showed better recognition memory for pictures which were presented after an event boundary than for pictures which were presented before an event boundary. This finding suggest that the surface information of recently encountered objects would become less available after the event model is updated. Swallow et al. (2009)'s study used film clips during which target objects was presented for a few second. This study found that participants' recognition memory for object was significantly affected by whether an even boundary occurred in the film clip during a 5s delay from the offset of the object presentation until the onset of the memory test.

A recent series of experiments investigated the effect of spatial shift on participants' recognition memory of objects presented in a virtual reality environment (Radvansky & Copeland, 2006; Radvansky, Krawietz, & Tamplin, 2011; Radvansky, Tamplin, & Krawietz, 2010). These studies found that participants were more likely to forget the recently encountered objects when they moved from one room to another, compared to when they stayed in the same room. The shift of location served as an event boundary and prompted people to update their current event model. After experiencing an event boundary, information linked to the prior event would become less accessible in one's memory. This 'location updating effect' was observed not only in a virtual reality environment, but also in a real-world setting where participants moved across real rooms (Radvansky et al., 2011). The location updating effect was further distinguished from the encoding specificity account (Thomson & Tulving, 1970) or the context-dependent memory. According to the encoding specificity account, if people return back to the original spatial context, their memory would be improved, because the original context provides more retrieval cues which facilitate the access of target information. Radvansky and his colleagues conducted one study to test this alternative explanation (Radvansky et al., 2011), but they found no improvement in participants' recognition memory of recently encountered objects by returning to the original context. This finding suggested that the cognitive mechanism under which event segmentation influences memory cannot be fully explained by the encoding specificity account.

### 4.4.3. Protagonist shift as event boundary

As introduced earlier in this chapter, protagonist is one of the five key dimensions that people rely on in order to index the event models. When people are engaged in communication, they construct event models based on the dimension of protagonist. Therefore, shifts in protagonists (e.g., conversational partners) would influence how the event models is constructed and how people's memory is structured. This protagonist shift effect could potentially explain how common ground information may appear to be partner-specific during communication and how this partner-specific effect can be explained by a relatively low-level memory mechanism. According to the event segmentation theory, two specific hypotheses can be derived. Firstly, a shift in protagonists could potentially introduce event boundaries during event processing, which cause the current event model to be updated. Secondly, a shift in protagonist could also have an effect on people's memory encoding and retrieval. Information represented in a protagonist-associated event model should become less available after a protagonist shift.

Most of the studies on protagonist shift provide strong evidence in support of the first hypothesis that character shifts do create event boundaries during reading processing (e.g., Magliano, Miller, & Zwaan, 2001; Rinck & Weber, 2003; Zwaan, Radvansky, Hilliard, & Curiel, 1998). Some studies examined participants' reading times of a specific sentence as an indication of the processing load for reading that sentence. Zwaan et al. (1998) tested all five dimensions proposed in the event-indexing model to investigate their relative importance in a constructed mental model. Results showed that sentences with discontinuities on temporal, protagonist, goal-intentional or causal dimension reliably increased people's reading times of those sentences. Reading times increased as a function of the number of situational shifts occurred. Rinck and Weber (2003) tested the effect of discontinuity in space, time and protagonist on the processing load of target sentences. Results showed that the protagonist shift did reliably increase participants' reading times of a target sentence where such a shift was introduced.

It seems compelling that protagonist is one of the basic dimensions that people depend on to construct event models. However, few studies have examined the effect of protagonist shifts on memory encoding and retrieval. One exception was a study by Rich and Taylor (2000), which examined the effects of shifts in protagonist (i.e., introduction of a new protagonist), shifts in place (i.e., movement to a new location), and shifts in time (i.e., the jump of narrative time). Each narrative has six memory recognition probes, each as a three-word phrase describing a specific event that either happened in the text or was thematically possible but did not happen. When a probe appeared, participants were asked to respond 'yes' if the event had occurred in the narrative and 'no' if it had not. There were three types of probes: cross-section probes, within-section probes, and false probes. The cross-section probes described a previous event which occurred prior to a sentence that contained a shift in protagonist, place or time. The within-section probe described a current event that participants were experiencing. The false probe described an event that did not occur in the narrative. Results suggested that participants responded more accurately and quickly for within-section probes than cross-section probes. Moreover, among the three types of shifts, protagonist shifts seemed to have the strongest effect on people's memory. Participants responded significantly more accurately and significantly more slowly to the probes when following protagonist shifts than following either location shifts or time shifts. Rich & Taylor (2000) claimed that the protagonist dimension seemed to be the most important index among the three event indices examined in their study, as protagonist shifts had the greatest effect on participants' memory compared to the location shifts and time shifts.

### 4.4.4. The current work on the effect of conversational partner change on memory

Studies on event segmentation theory suggest that people keep track of protagonist-related information during an event, and that change of protagonist acts as event boundaries during event processing and results in the update of the current event model. Research on partner-specificity in referential communication suggests that speakers and listeners both keep track of partner-specific information in order to communicate successfully, and that the common ground status between a speaker and a listener will change if their conversation partner changes. Both lines of research focus on the change of person involved in an event and examine its effect on memory encoding and retrieval.

By combining the event segmentation theory with the memory-based model of conversational common ground, I propose that the event segmentation theory may offer an efficient way for partner-associated common ground to be encoded and stored separately in episodic memory traces. Changes of communication partners would serve as event boundaries, which divide on-going social activities and conversations into several event models. Each constructed event model is indexed by a specific communication partner (although other indices also exist). The specific partner serves as an informative cue during later memory retrieval, enabling the partner-associated common ground to be easily tracked during communication.

Studies in Chapter 5 tested whether event segmentation mechanism could potentially shape and support the memory representations of partner-specific information about common ground during online social interaction, and whether partner-associated common ground information can be established through such a low-level memory mechanism, without necessarily making deliberate Theory of Mind inference.

## 4.5. Influence of conversational partner's perspective status on memory

When we are actively involved in the process of perspective-taking during communication, we pay attention to our conversational partner's perspective to assess what is shared and what is not shared between us. Experiments in Chapter 6 introduced a novel approach to examine whether participants' memory for objects in a post-test was affected by whether object being encoded from the common ground or privileged ground during a preceding referential communication task.

### 4.5.1. Common ground versus privileged ground

The ability to establish and retain common ground information is essential for human communication. Common ground information is believed to be a foundation upon which people build everyday social lives through communication and interaction with others. The traditional view of cooperative communication (see '*cooperative principle'* discussed in Chapter 1), which was first proposed by Grice (1957), expect people to use common ground information from the earliest moments of language processing while completely ignoring information in privileged ground. However, empirical studies have demonstrated that although common ground information shapes what and how people communicate, communication is never free from the influence of information in privileged ground. Studies have shown that information in one's privileged perspective has an effect on language comprehension and production, either as a probabilistic constraint competing with common ground information (e.g., Brown-Schmidt et al., 2008; Hanna & Tanenhaus, 2004; Hanna et al., 2003; Nadig & Sedivy, 2002), or as an immediate constraint that needs to be corrected and adjusted later by drawing upon common ground information (e.g., Barr, 2008; Epley, Keysar, et al., 2004; Keysar et al., 2000, 2003).

Empirically, information in common ground and privileged ground both influences how people process languages (e.g., Keysar et al., 2000, 2003). Theoretically, however, common ground information is still considered to be prioritised over privilege ground information (e.g., Clark & Marshall, 1978, 1981). Consideration of common ground in communication is often related to the concepts of 'audience design', 'perspective taking' and 'partner-specific adaptation'; whereas attention to privileged ground, on the contrary, is often related to the concepts of 'egocentrism', 'failure of perspective taking' and 'miscommunication'. When describing results from communication tasks, researchers often use the term '*successfully* take into account the information in common ground' and '*fail to* completely ignore information in privileged ground' (e.g., Hanna et al., 2003). The use of these terms implies that the effect of common ground is desirable whereas the effect of privileged ground is unwanted for language comprehension and production. Similarly, when analysing participants' eye movements data in the director task, numbers of fixations and duration of looking time on the distractors in the privileged ground are often considered to reflect participants' degrees of egocentric bias (Keysar et al., 2000, 2003; Lin et al., 2010; Nilsen & Graham, 2009).

Recent research has only started to examine the benefits of privileged ground information for communication. A study by Brown-Schmidt et al. (2008) shifted the research focus from the use of common ground information to the use of both common ground and privileged ground information. This study found that when speakers asked questions to inquire about information only visible to listeners ('what's above the cow with shoes?'), listeners rapidly directed their attention toward items in their privileged ground even before hearing the full sentence. Conversational participants was found to be sensitive to the forms of utterances. Specifically, the use of questions directed participants' attention to privileged ground information, whereas the use of imperatives directed participants' attention to common ground information. Therefore, Brown-Schmidt et al. (2008) proposed that 'selective attention to privileged ground when being asked for information may reflect awareness of the knowledge states of others, rather than egocentricity' (p1123).

The use of common ground information is necessary but insufficient for one's success in taking a conversational partner's perspective. Successful communication requires a sophisticated process of balancing the information from both the common ground and the privileged ground in a context-specific way. Different communicative contexts may require people to pay more attention to information in the common ground or privileged ground.

### 4.5.2. The current work on the effect of perspective status on memory in Chapter 6

When we are engaged in a conversation, how do we distribute our attention to information in the common ground and privileged ground? When we are required to take a conversational partner's perspective into account, do we spontaneously pay more attention to information in common ground compared to privileged ground? Does the ground status of objects (i.e., common ground vs. privileged ground) influences the representations of these objects in memory? Do our attention and memory systems grant common ground information a prioritised status overs privileged ground information?

Studies in Chapter 6 took a memory-based approach to address these questions regarding the attention directed to information in common ground versus privileged ground, and the memory representations of common ground versus privileged ground. If more attention is assigned to common ground objects during perspective-taking, then participants are expected to have better memories for objects in the common ground than those in the privileged ground in a subsequent object recognition test.

## 4.6. Overview of Chapters 5 and 6

Chapters 5 and 6 took a memory-based approach to study how common ground information may be encoded and stored in memory, and how people access the encoded common ground representations. Experiments were designed to examine whether people's memory for objects can be influenced automatically by common-ground-related social contexts (e.g., whether an object was probed in the presence of the same / different person as it was encoded, or whether an object was visible / in visible to another person). If the common-ground-related social contexts have an effect on people's memories for objects even when this effect is not required or even undesirable, then it is likely that common ground representations can be established under a relatively low-level memory mechanism which requires little explicit inference and cognitive effort.

Chapter 5 asked whether a change in the identity of a communicative partner influences participants' memories for objects associated with different communicative partners. I propose that people may draw upon the associative memory traces to keep track of partner-specific representations in a domain-general manner. Chapter 6 examined whether the status of an object in a conversational partner's perspective (i.e., common ground versus privileged ground) would spontaneously influence participants memories for the object. Participants were expected to show a better recognition memory for objects in the common ground than those in the privileged ground.

CHAPTER V

# Is memory for objects influenced by their being associated with a previous or a current conversational partner?

## 5.1. Introduction

In Chapter 4, I reviewed evidence of how people might assess common ground by drawing upon a low-level memory-based mechanisms and how event segmentation mechanism might automatically structures our memory into meaningful chunks associated with partner information. By combining the low-level memory-based model of common ground assessment (Horton & Gerrig, 2005a, 2005b) with the event segmentation theory (e.g., Kurby & Zacks, 2008; Zacks & Swallow, 2007), Chapter 5 aimed to explore whether the change of conversational partners might act as an event boundary and divide on-going communication into different event models, which offers an efficient way for common ground information to be encoded in separate episodic memory traces which are associated with specific partners. If such partner-specific associations was established and stored in memory, then the specific partner was expected to serve as an informative cue during later memory retrieval, enabling the partner-associated common ground information to be easily accessed during communication. Experiments in Chapter 5 was designed to investigate whether common ground assessment might be achieved during conversation via such a low-level memory-based mechanism, without necessarily making deliberate Theory of Mind inference.

Firstly, experiments in Chapter 5 gained inspiration from a recent study by Samson et al. (2010) on automatic and spontaneous perspective taking. Samson et al. (2010) found that an avatar's visual perspective affected participants' own judgement about what they see, even when participants were only asked to judge their own visual perspective but never the avatar's perspective. Samson et al. (2010) employed such a paradigm where the avatar's perspective was irrelevant to participants' task and was thus unlikely to be the object of strategic attention, but they still observed a marked impact of the avatar's perspective on participants' performance. They therefore concluded that people compute other's perspectives into account in a relatively automatic fashion. In the current study, we took a similar approach to examine whether the availability of objects in memory would be affect by whether or not the conversational partner associated with the specific object was present at the point of memory retrieval. If such a partner-specific effect emerged even when it was detrimental to participants' performance, and when participants were not explicitly asked to bind the conversational partner information with the associated objects, then the finding would provide evidence that partner-specific effect can be established by a relatively low-level memory mechanism.

Secondly, the present experimental design also draws inspiration from a series of studies on event cognition by Radvansky and his colleagues (Radvansky & Copeland, 2006; Radvansky et al., 2011, 2010). In the studies, participants were instructed to move from one room to another room in a virtual environment containing several rooms, where each room had one or two tables. Participants were asked to first pick up one object from a table, then walk to another table carrying the object, set the object down and then pick up another object. At critical points during the task, participants were probed with object images, and were asked to indicate whether they recognised the object as one of the two recent objects they interacted with. The memory test occurred either upon participants' entrance into a new room or at the halfway point across a large room. Results showed that when people moved from one room to another, they were more prone to forgetting information than when they did not make such a spatial shift. Radvansky and his colleagues suggested that moving from one room to another (i.e., spatial shift) serves as an event boundary, which prompted people to update their current event model. The day-to-day activities in which people engage are parsed into different event models in this way. Information associated with the prior event becomes less accessible in memory after the event boundary. Information within each event model shares strong associations, while information between different event models shares relatively weak associations. This event segmentation mechanism helps people to effectively organise the encoded information into memory chunks as they go through different events, which also benefit later memory retrieval.

Event segmentation theory suggests that the change of protagonists, as well as the change of locations, prompts an event boundary which shapes the memory structure. Therefore, we designed a computerised 'shopping task' where the manipulation of the change of conversational partners was innovatively implemented. In the 'shopping task', participants interacted with a series of objects introduced by various agents. Each agent introduced two objects to participants, one at a time. Frequent object-recognition probes appeared either after a change of agents or no change of agents. The change of agent-identity was manipulated to investigate whether changes of communication partners would make the most recently encoded two objects less accessible in memory. We expected the event-model-updating effect to be observed in the present experiment. Specifically, people would tend to show more errors and slower responses when they were probed in the presence of a new agent than in the presence of the previous agent. If our hypothesis is confirmed, then this may provide insights about how partner-specific common ground may be achieved implicitly, efficiently and automatically under such a low-level memory mechanism shaped by the way we construct event models.

## 5.2. Experiment 4

### 5.2.1. Methods

#### *5.2.1.1. Participants*

Twenty-four students (20 females, mean age 20.83, age range 18 to 32) from the University of Birmingham participated in this study in return for a small honorarium or course credits. All participants reported normal colour vision, and normal or corrected-to-normal visual acuity. Ten additional participants were excluded prior to the analysis due to failure to perform above chance level.

#### *5.2.1.2. Materials and apparatus*

Fifty-six agent images were created with an iPad App named '3D Agent Creator'. Each agent had distinct physical features (e.g., eye shape, complexion, hair-styles and clothing.) Examples of the agent images employed in this study are provided in Figure 5.1. One shared feature among all agents was that they all contained exaggerated eye-features. One study suggested that presenting cartoon-like agent images with salient eye-features rather than human-like agents with realistic eye-features helped participants to focus more on agents' eye gaze (Conty, Tijus, Hugueville, Coelho, & George, 2006). Therefore we expected that by presenting agents with exaggerated eye features, participants' attention would be directed more towards the agents than other places. Fifty-two agents were presented in the experimental phase, while another 4 agents were presented in the practice phase. All the objects employed in our study were created by combining 10 colours (i.e., red, orange, yellow, green, blue, purple, white, grey, brown and pink) and 13 everyday objects (e.g., bag, hat, clock, cup, glassed, book, pants, scarf, shirt, shoes, socks, tie and umbrella).

The experiment was visually presented on a 15.6-inch Samsung laptop. E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used for the presentation of all materials and for recording participants' response time and accuracy.

Figure 5.1. Examples of agent images used in Experiment 4.

#### *5.2.1.3. Design and procedure*

**The Shopping Task**

The computerised '***shopping task'*** was designed for the present study. Fifty-two agents were presented to participants sequentially. Each agent introduced two objects to participants, one object at a time. Participants were asked to put each object into their shopping basket using a computer mouse, and then remove it out of basket before putting in another object. On 48 out of 104 trials a probe image would appear after one object had just been put into the basket, either during the presence of the same agent who was associated the object (i.e., the no-change condition) or during the presence of a new agent (i.e., the change condition). The participants were instructed to respond 'yes' by pressing the left button on a computer mouse if the probe was either the object they had just put into the shopping basket (the current object) or the object they had just taken out of the basket (the previous object). Figure 5.2 illustrated the schematic procedure of the 'shopping task' for trials which presented current-object probes (which was referred to as 'current-object trials' in the following sections), and

Figure 5.3 for trials which presented previous-object probes (i.e., 'previous-object trials'). Participants were asked to respond 'no' by pressing the right button if a probe was an mismatched objects that recombined shape and colour features from the current and previous objects (i.e., foil probes). It was equally likely for a foil to be of the colour of the previous object and the shape of the current object versus the shape of the previous object and the colour of the current object. Each foil appeared only once, and was never used as positive probes. Participants' response time and accuracy were recorded.

In the current study, participants were asked to hold in mind both the current object and the previous object at the same time and to update their mental representations of the recent two objects every time when a new object was introduced, instead of holding in mind only the current object. This design choice was made for three reasons: 1) to increase the difficulty level of the memory task; 2) to create foils by mismatching the shape and colour features from the current and previous objects; 3) to implement the change of conversational partners into a similar experimental paradigm which has already been proven to be able to detect event updating effects in Radvansky and his colleagues' studies (Radvansky & Copeland, 2006; Radvansky et al., 2011, 2010), whilst minimizing additional adaptations as our first attempt[7].

We were primarily interested in the effect of conversational partner change on memory. Our predicted patterns of results were: 1) Participants would show more accurate memory for each type of memory probes when they were probed in the no-change condition than in the change condition; 2) Participants would take shorter time to make a correct response to each type of memory probes when they were probed in the no-change condition than in the change

[7] In most of Radvansky and Colleagues' computerised experiments, participants were also required to remember and recognise the two most recently interacted objects (Radvansky & Copeland, 2006; Radvansky, Krawietz, & Tamplin, 2011; Radvansky, Tamplin, & Krawietz, 2010).

condition. In accordance with the findings from Radvansky and colleagues, we also expected that participants' memories for current objects would be better than their memories for previous objects.

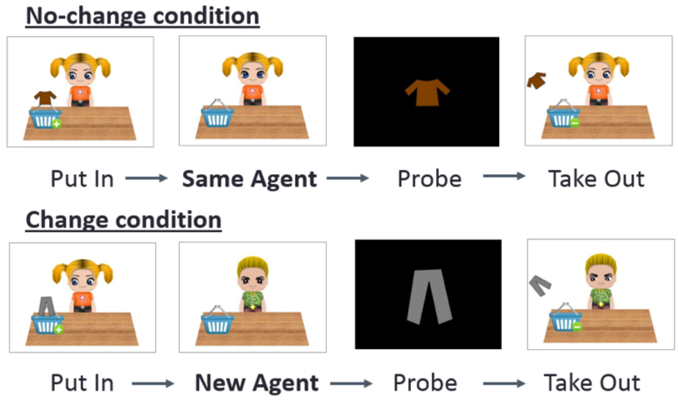


Figure 5.2. Schematic presentations of current-object trials in the 'shopping task'.

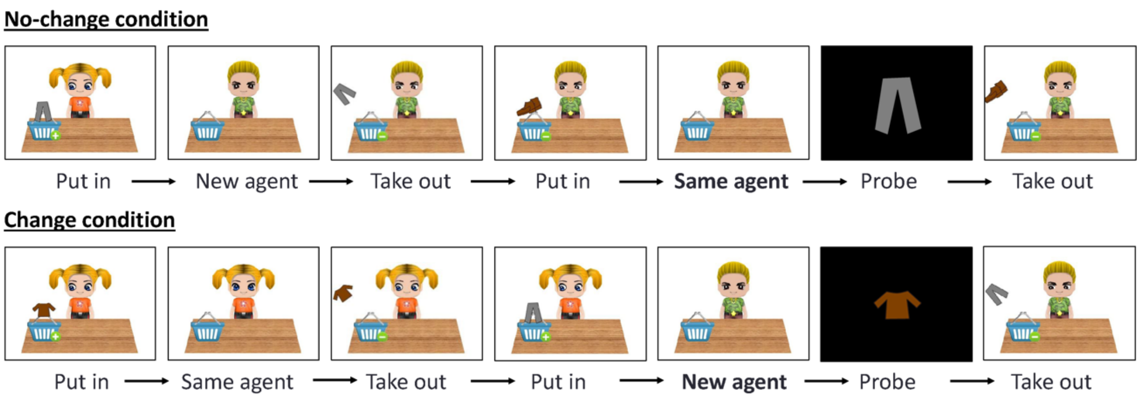


Figure 5.3. Schematic presentation of previous-object trials in the 'shopping task'.

**The within-participants design**

A within-participants design was constructed with condition (change and no-change conditions) and probe-type (positive probes and foils) as two independent variables. Within the positive probes, there were two sub-types: the current-object probes and the previous-object probes. The structure of this within-participant experimental design is illustrated in Figure 5.4. The whole experiment consisted of 1 practice block (8 trials) and 4 test blocks (26 trials each). Participants were allowed to take breaks between blocks. During the test phase, each participant completed 104 trials in total, 48 of which were accompanied by a memory probe. Not every possible test point was associated with a memory probe. This was in order to make it less likely for participants to be able to predict where probes would appear. There were 24 probe trials for the change and no-change conditions respectively, resulting in 6 trials presenting current-object probes, 6 trials presenting previous-object probes, and 12 trials presenting foils in each condition. The order of blocks was counterbalanced across subjects. The order of trials within each block was fixed based on a pseudo-random assignment of different conditions and probe types, with the criteria being that (a) participants never encountered more than three probe trials continuously, (b) participants never encountered three or more probe trials from the same condition in a row, (c) and the trial immediately before a trial presenting a previous-object probe never contained a memory probe.

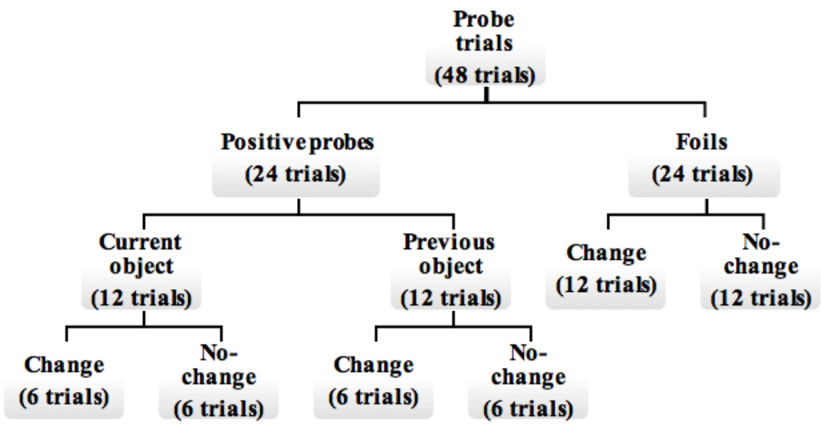


Figure 5.4. The structure of the within-participant experimental design in Experiment 4.

**Test Procedure**

After signing an informed consent form, participants were seated in front of a laptop. They were instructed to imagine that they were shopping in a supermarket, and to regard the agents on the screen as shop assistants. At the beginning of each trial, an agent was standing behind a table and introducing an object placed on the right side of the table by a series of eye gaze shifts and hand gesture. Research has suggested that non-verbal communicative cues such as eye gazes and pointing gestures can direct people's attention towards the target stimuli (Bristow, Rees, & Frith, 2007; Langton, Watt, & Bruce, 2000). Then a shopping basket with a plus sign (+) appeared on the left side of the table. At this point, the agent averted his/her eye gaze towards the basket, indicating participants to put the object into the basket. Participants were asked to drag the object into the shopping basket using a computer mouse. The object was no longer visible to participants as soon as it was put into the basket. Before introducing participants to another object, the agent again averted his/her eye gaze towards the basket. Participants were instructed to take the object out of the basket by clicking on a minus sign (-) on the basket. When the minus sign was clicked, the object flied out of the basket and disappeared from the left side of the screen (which lasted for 600ms). After one agent introduced two objects in sequence, a second agent introduced another two objects, and so forth. The change of agents is displayed by the previous agent fading out gradually (900ms) and then the new agent fading in (900ms). In order to match the duration of trials, in the no-change condition, the existing agent was displayed on the screen for the same duration (1800ms). In the change condition, the memory test appeared after a new agent has faded in; while in the no-change condition, it appeared after the most recent product had been put into the shopping basket and before the same agent presented another product. In the memory test, a white cross was first presented in the centre of a black screen for 1000ms, followed by a probe image. The probe image stayed on the screen until participants made a response. Figure 5.5 illustrates an example of the event sequence of a previous-object trial in the change condition, and Figure 5.6 illustrates an example of the event sequence of a current-object trial in the no-change condition. The experimental procedure typically lasted for 25 to 35 minutes. Participants were fully debriefed at the end of the experiment.

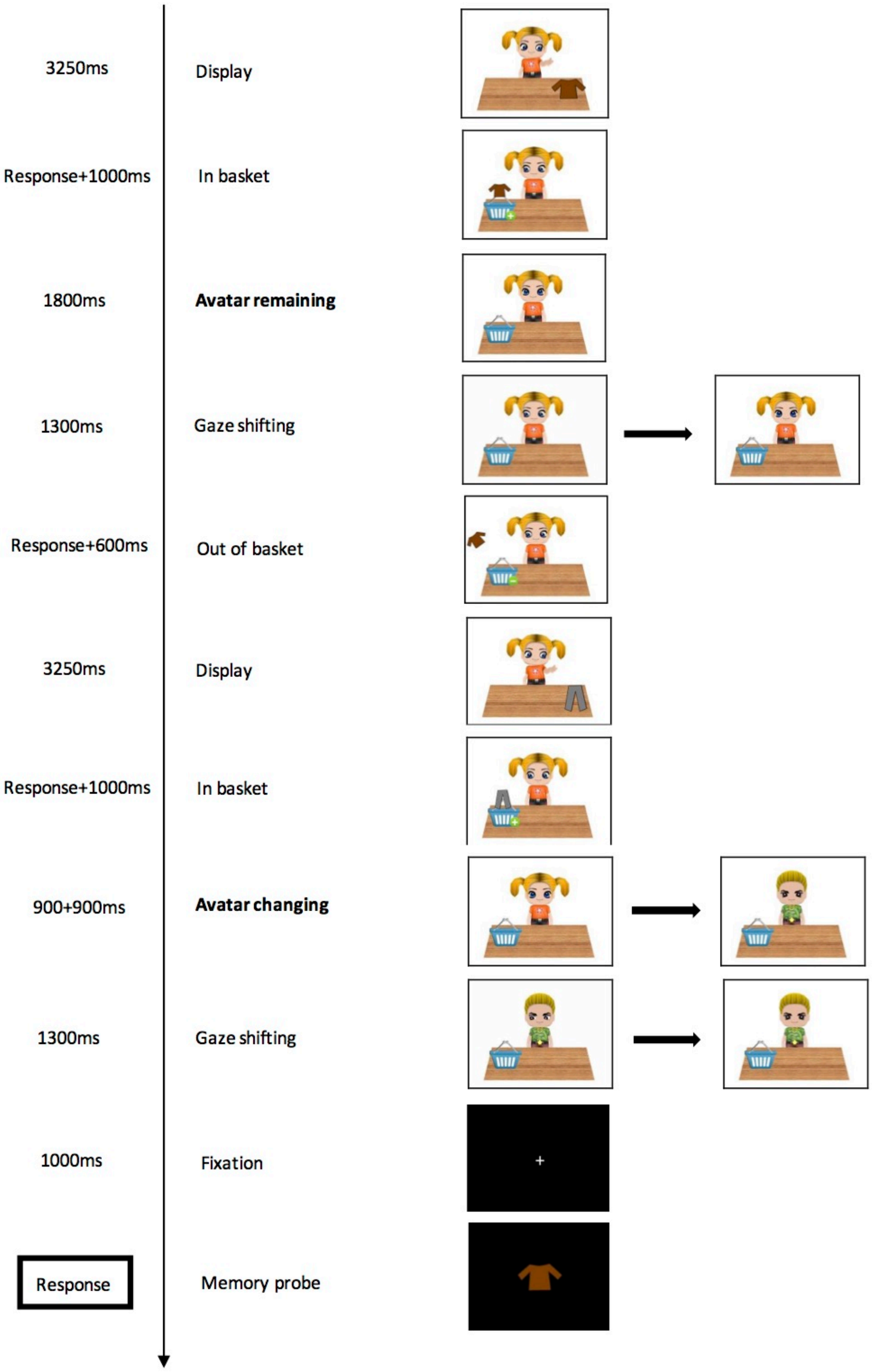


Figure 5.5. Schematic diagram of the event sequence in a previous-object trial in the change condition.

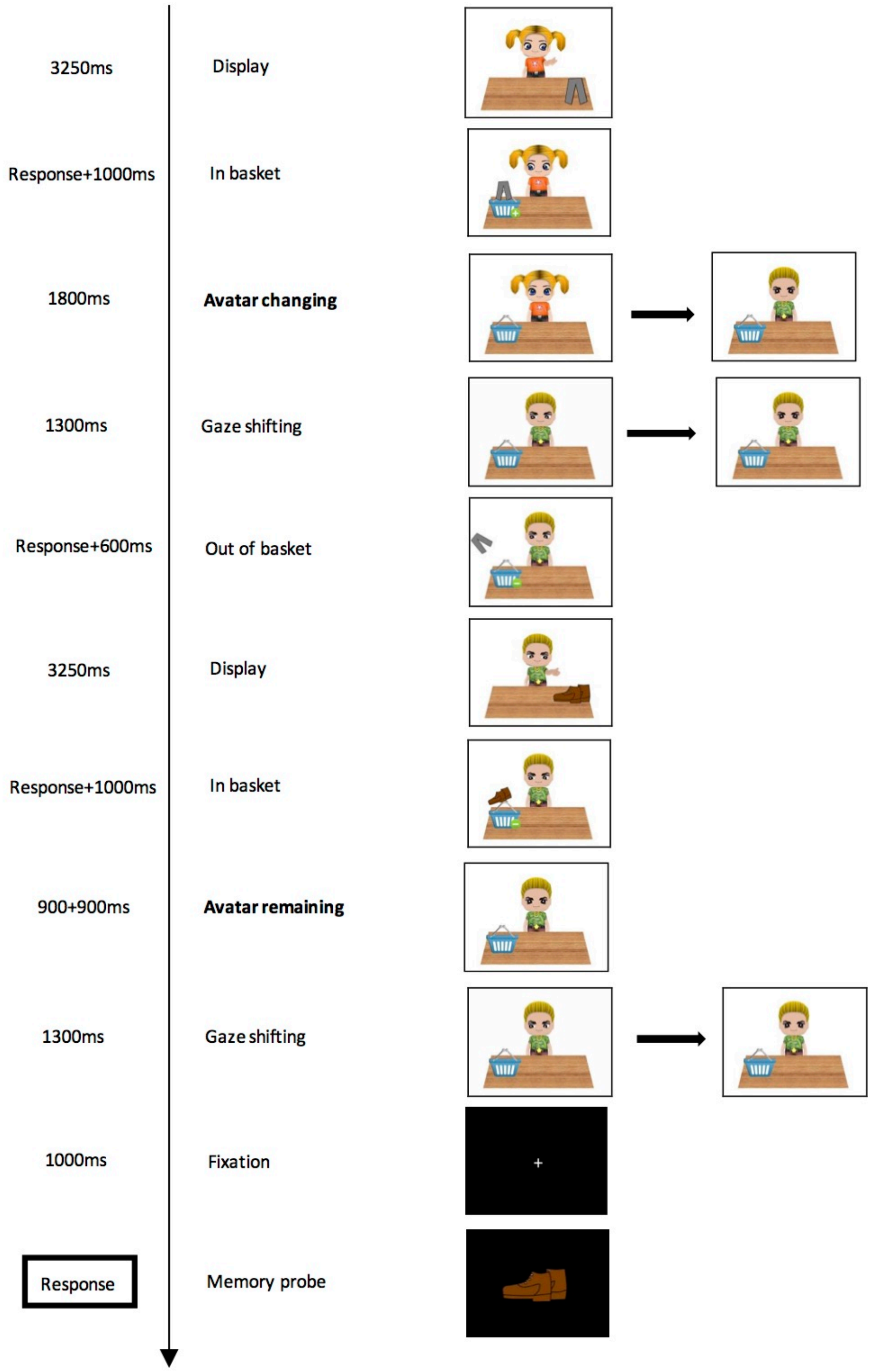


Figure 5.6. Schematic diagram of the event sequence in a current-object trial in the no-change condition.

### 5.2.2. Results

Twenty-four participants whose accuracy was significantly above chance level for all three probe-types (current-object probes, previous-object probes and foil probes) were included in the final analysis.

Accuracy and response time data were analysed separately (see Table 5.1). The response time analysis only included trials to which participants responded correctly. Response times longer than 2.5 standard deviations from mean response time were also excluded prior to the analysis, which caused 1%- 2% (varying with probes types) of the response time data from correctly responded trials to be dropped. Responses quicker than 200ms were generally considered as abnormal  by many studies (e.g., Ballesteros et al., 2015; Grubert & Eimer, 2016; Hübner & Többel, 2012; Lupianez, Ruz, Funes, & Milliken, 2007) and hence were planned to be excluded prior to our analysis. In the current data, however, no response was found to be quicker than 200ms.

Table 5.1. Accuracy (proportions) and Response Time (ms) with standard deviations for Experiment 4.

| | **Accuracy** | | | | **Response time** | | | |
|---|---|---|---|---|---|---|---|---|
| | **No-change** | | **Change** | | **No-change** | | **Change** | |
| | M | SD | M | SD | M | SD | M | SD |
| **Current objects** | 0.95 | 0.09 | 0.88 | 0.13 | 779 | 137 | 826 | 181 |
| **Previous objects** | 0.92 | 0.13 | 0.89 | 0.13 | 791 | 138 | 895 | 226 |
| **Foils** | 0.89 | 0.08 | 0.91 | 0.10 | 945 | 231 | 1021 | 245 |

As the trials presenting positive probes and foil probes involved different types of responses, they were analysed separately. For trials which presented positive probes, 2 x 2 repeated-measures ANOVAs were conducted with condition (no-change vs. change) and probe-type (current vs. previous object) as the main contrasts, for accuracy and response time respectively. For trials which presented foil probes, the accuracy and response time data were submitted to paired t-tests with the change/no-change condition being a within-participants factor. If participants showed significantly stronger memory traces (reflected by higher accuracy or quicker response) for each probe type in the no-change condition than in the change condition, this would provide evidence in support of our hypothesis that participants may form associations between objects and the agents in memory automatically.

We first reported the ANOVA results for trials which presented positive probes. For accuracy, there was a marginally significant main effect of condition, $F(1, 23) = 4.054$, $p = .056$, $\eta p^2 = .150$. Participants were marginally significantly more accurate in the no-change condition than the change condition. There was no main effect of probe type, $F(1, 23) = .294$, $p = .593$, $\eta p^2 = .013$, and also no interaction of probe type x condition, $F(1, 23) = .869$, $p = .361$, $\eta p^2 = .036$. Planned comparisons were carried out to examine the main effect of change/on-change condition, for current-object trials and previous-object trials separately. For current-object trials, participants made significantly more errors when a change of agents happened than when no such change happened, $t(23) = 2.105$, $p = .046$. For previous-object trials, participants tended to be less accurate in the change condition compared to the no-change condition on average, but there was no statistically significant difference, $t(23) = .893$, $p = .381$.

For the response time data, a significant main effect of change/no-change condition was found, $F(1, 23) = 8.864$, $p = .007$, $\eta p^2 = .278$. Participants spent significantly longer time to make a correct response when they were in the change condition compared to the no-change condition. There was main effect of probe types, $F(1, 23) = 1.359$, $p = .256$, $\eta p^2 = .056$, and no significant interaction between probe types and conditions either, $F(1, 23) = 1.601$, $p = .218$, $\eta p^2 = .065$. The data from current-object and previous-object trials were then entered into paired t-tests respectively to examine the change/no-change effect. For current-object trials, although participants' response time to memory tasks was slightly longer in the change condition than in the no-change condition, the difference was not statistically significant, $t(23) = 1.422$, $p = .168$. For previous-object trials, it took participants significantly longer to make correct responses in the change condition than in the no-change condition, $t(23) = 2.584$, $p = .007$.

For trials involving foil probes, the accuracy and response time data were submitted to paired t-tests to make comparisons between change and no-change condition. For accuracy, there was no significant difference between conditions, $t(23) = .862$, $p = .397$. However, the response time was significantly longer in change condition than in no-change condition, $t(23) = 2.584$, $p = .017$.

### 5.2.3. Discussion

The results of Experiment 4 generally confirmed our hypothesis. Participants made significantly more errors and longer responses in memory tasks if there was a change of communication partners between the memory encoding point and retrieving point. This experiment showed a significant effect of agent change. Although the information regarding agent change was irrelevant to the memory task and participants did not need to pay attention to this information, their performance was still influenced by whether or not there was a change of agents.

Nonetheless, one alternative explanation could not be dismissed based on the present experimental design. In the change condition, the change of agents was displayed by the previous agent fading out from the computer screen (900ms) and then the new agent fading in on the screen (900ms). In contrast, in the no-change condition the existing agent was displayed on the computer screen continuously for the same total duration (1800ms). There was thus a possibility that participants showed a declined memory in the change condition merely because of the visual distraction caused by a previous agent gradually fading out and new agent fading in, rather than the change of agents' identity per se. Experiment 5 was designed to examine this alternative explanation.

## 5.3. Experiment 5

The aim of Experiment 5 was to test whether the significant effect found in Experiment 4 resulted from the change of agent-identity, or alternatively from the visual distraction of a previous agent fading out and a new agent fading in in the change condition. In Experiment 5, both the change and no-change conditions contained sequences of agents coming in and out of a scene. Therefore, the effect of agent change on participants' recognition memory was observed again in Experiment 5, then we could reject the suggestion that declined memory in the change condition was merely resulted from visual distraction of agents fading in and out of the display.

### 5.3.1. Methods

#### *5.3.1.1. Participants*

Thirty-one participants (25 females, mean age 20.71, age ranges from 18 to 38) from the University of Birmingham participated in this study in return for a small honorarium or course credits. All participants reported normal colour vision, and normal or corrected-to-normal visual acuity. Twelve additional participants were excluded prior to analysis due to failure to perform above chance level.

#### *5.3.1.2. Materials and apparatus*

The materials and apparatus used in Experiment 5 were identical to those used in Experiment 4.

#### *5.3.1.3. Design and procedure*

The experimental design and procedure in Experiment 5 were identical to those in Experiment 4 with the following exception. In Experiment 5, an image of a wooden door was added to the upper right corner of the computer screen during experiment. In the change condition, after each introduced object was put into the basket by the participants, the previous agent left through the door (this sequence lasted for 1850ms), and then the new agent came into the room through the door (1850ms). In the no-change condition, the present agent left the room through the door (1850ms), and then the same agent came back in again (1850ms).

### 5.3.2. Results

The descriptive results of the accuracy and response time data are summarized in Table 5.2. The same protocol we employed to clean and analyse data in Experiment 4 was applied to data from Experiment 5.

For trial which presented positive probes, a 2 x 2 repeated-measures ANOVA was conducted on the measure of accuracy and response time separately, with condition (no-change vs. change) and probe-type (current vs. previous object) as the within-subject factors. For accuracy, there was no main effect of condition, $F(1, 30) = 1.972$, $p = .170$, $\eta p^2 = .062$, no main effect of probe type, $F(1, 30) = 2.154$, $p = .153$, $\eta p^2 = .067$, and a marginally significant interaction between probe types and conditions, $F(1, 30) = 3.433$, $p = .074$, $\eta p^2 = .103$. Planned comparisons were carried out to examine the main effect of the change/on-change condition, for the current-object and previous-object trials separately. For current-object trials, there was no difference between the change condition and no-change condition, $t(30) = .407$, $p = .687$. Strikingly, for previous-object trials, participants made significantly fewer errors in the change condition than in the no-change condition, $t(30) = 2.094$, $p = .045$, contrary to our prediction.

For response time, there was no main effect of condition, $F(1, 30) = 2.048$, $p = .163$, $\eta p^2 = .064$, and no main effect of probe type, $F(1, 30) = .093$, $p = .763$, $\eta p^2 = .025$. However, there was a significant interaction between probe types and conditions, $F(1, 30) = 4.528$, $p = .042$, $\eta p^2 = .131$. The data from the current-object and previous-object trials were entered into planned t-tests respectively to make comparisons between the change and no-change condition. For current objects, participants' responses were significantly slower in the change condition than in the no-change condition, $t(30) = 2.445$, $p = .021$. For previous objects, there was no difference between the two conditions, $t(30) = .150$, $p = .882$.

For trials which presented foil probes, the accuracy and response time data were submitted to paired t-tests with the change/no-change condition as a within-subject factor. For accuracy, participants showed no difference between the change condition and no-change condition, $t(30) = .822$, $p = .417$. For response time, participants' response on average tended to be slower in the change condition than in the no-change condition, however this difference failed to reach significance, $t(30) = 1.692$, $p = .101$.

Table 5.2. Accuracy (proportions) and Response Time (ms) with standard deviations for Experiment 5.

| | **Accuracy** | | | | **Response Time** | | | |
|---|---|---|---|---|---|---|---|---|
| | No-change | | Change | | No-change | | Change | |
| | M | SD | M | SD | M | SD | M | SD |
| **Current objects** | 0.95 | 0.09 | 0.93 | 0.11 | 910 | 280 | 1019 | 345 |
| **Previous objects** | 0.88 | 0.14 | 0.95 | 0.11 | 983 | 478 | 977 | 362 |
| **Foils** | 0.94 | 0.09 | 0.92 | 0.10 | 1059 | 334 | 1117 | 331 |

## 5.3.3. Discussion

Experiment 5 was conducted to test the alternative explanation that the significant effect of conversational partner change in Experiment 4 was merely a consequence of the distractive fading effect, which occurred in the change condition but not in the no-change condition. In Experiment 5, the sequences in the change and no-change conditions were closely matched. Results of Experiment 5 depicted a rather mixed picture. Contrary to our expectation, the accuracy result for the previous-object trials showing weaker memory trace for previous objects in the no-change condition compared to the change condition. However, in favour of our hypothesis, participants took significantly longer time to make a correct response in the change condition compared to the no-change condition on the current-object trials, and also marginally significantly longer time on the trials which presented foil probes. Findings from Experiment 5, therefore, failed to rule out the alternative explanation that the visual distraction of agent-fading caused declined memory in the change condition compared to the no-change condition.

## 5.4. Discussion of Experiments 4-5

Experiments 4 and 5 was intended to demonstrate that changing communication partners in a social situation reduces the availability of recently encoded objects in memory. Two experiments were carried out using a computerised 'shopping task', during which the change of agent-identity was manipulated and participants were asked to recognise three types of objects (current, previous and foils) in either the change or no-change condition. Results in Experiment 4 were largely consistent with our hypothesis, showing declined memory after a change of agents. Findings from Experiment 5 showed a mixed pattern: the significant effect observed for current object probes and foils probes in Experiment 5 seemed to indicate that the change of agents indeed had a negative effect on the accessibility of recently encoded objects in memory, but the results for previous object probes did not support our hypothesis.

The mixed pattern of results from Experiment 5 drove us to have a closer look at the trial structure for each type of objects. The manipulation of the change/no-change condition was straight-forward for the current object trials. The only difference between these two conditions was whether a previous agent was replaced by a new agent before a memory probe appeared. However, the structure of the trials which presented previous-object probes was relatively complicated. Participants could see and encode the previous object twice: the first time when participant put the object into the basket, and the second time when they took the object out of basket. In Experiments 4 and 5, the label 'change/no-change condition' was defined as whether participants encountered a new agent right before the memory test. This label may appear to be confusing for previous-no-change trails, as it also contained a change of agents—although such a change did not occur right before the memory test, it happened between the first and the second opportunity to encode information. Since a reversed pattern of agent change effect was observed on previous-object trials in Experiment 5, it was unclear whether such a complex experimental design might have obscured the predicted effect of agent change on memory. Follow-up experiments 6 and 7 were therefore designed to eliminate this potential confounding effect, by removing the previous-objects probes and increase the number of current-object probes in one memory test. An additional concern was that each participants only completed six experimental trials for either the current or previous object type in the change condition and in the no-change condition respectively, which may have caused the lack of expected effects in Experiment 4 and 5. Follow-up experiments should be designed to increase the data points collected from each participant in each experiment condition in order to enhance the possibility of detecting the predicted effects.

One issue in the current study is the high exclusion rate in the data. Ten out of thirty-four participants in Experiment 4 and twelve out of forty-three participants in Experiment 5 were excluded prior to analysis. Participants whose accuracy was not significantly above chance level in any of the three probe-type subgroups were exclude from analysis, because it was likely that their average or below-average performance resulted from the failure to follow the instructions of the experiment. The 'shopping task' required participants to hold in mind two objects (i.e., a previous object and a current object) at the same time, and more importantly, to update one of these two objects every time a new object was introduced. This instruction might be confusing for some participants, thereby causing the relatively high exclusion rate. Nonetheless, the instructions we used were similar to the instructions used in Radvansky et al.'s study on spatial updating[8], which did not indicate that participants had any difficulty in following instructions. Nonetheless, in the following experiments, effort was made to develop more appropriate instructions in order to reduce the exclusion rate.

Moreover, the memory test in the current design consisted of a sequence of single recognition probes presented at the centre of a full-screen black background. Such a setting of the memory phase was completely different from the setting of the presentation phase, and this sudden visual change may have created an unintended event boundary, which might have affected participants' performance. Another concern is that having the same agent leaving and returning the room in the no-change condition might have also introduced an unintended event boundary, as participants might have started to update the current event model upon the agent leaving the room. Future experiments should try to eliminate the possibility that a new event boundary is to be inserted by the change of settings between the presentation phase and test phase, and by the potential interruption of the same agent leaving and returning the room in the no-change condition.

[8] In most of Radvansky et al.'s experiments (Radvansky & Copeland, 2006; Radvansky et al., 2011, 2010), participants were required to put down an old object and pick up a new object while moving across the space. Participants were asked to recognise whether the memory probe was either the object that was currently being carried or the one that had just been set down.

## 5.5. Rationale for Experiments 6-7

Experiments 6 and 7 continued to address whether participants could automatically form associations in memory between objects and conversational partners. We expected participants' memory of a recently-encoded object to be less available when the object was probed in the presence of a new conversational partner than in the presence of the same partner. A new experimental paradigm was used to eliminate potential problems with the paradigm of Experiments 4 and 5 which may have yielded the inconsistent findings.

*Firstly*, the new experimental paradigm adapted the premise of the 'director task' paradigm and participants were asked to follow several directors' verbal instructions to select objects. In Experiments 4 and 5, although we had incorporated social cues such as eye gazes and hand gestures in order to emphasize the social communicative element of the 'shopping task', the interaction between participants and the agents was still limited. By using a communication task, Experiments 6 and 7 incorporated one-way verbal interaction between the agents and participants to enhance the social communicative context of the study. Each agent also had a unique voice in order to strengthen the identity of the agents. A small pool of agents was presented repeatedly in Experiments 6 and 7, with the aim to overcome potential visual distraction of introducing brand new faces on each experimental trial in Experiments 4 and 5.

*Secondly*, a group of 6 objects was presented at the same time and also occluded at the same time to make sure that each object was shown to participants for the same duration. On each trial, participants was asked to remember a group of 6 objects altogether. In Experiments 6 and 7, there was no such distinction between current and previous objects.

*Thirdly*, the memory test took place in the presence of an agent, rather than over a black background, in order to eliminate the possibility that a new event boundary might be inserted by the change of settings between the presentation phase and test phase. Memory probes consisted of true probes and foils. The foils were made of new colours and shapes (different from any of the colours or shapes presented on the same experimental trial), rather than the mismatch of the shape and colour of the current and previous objects in Experiments 4 and 5, Therefore, each colour and shape would only be presented once during each memory test.

## 5.6. Experiment 6

### 5.6.1. Method

#### *5.6.1.1. Participants*

Thirty students (2 males, mean age 18.9, age range 18 to 21) from the University of Birmingham participated in this study in return for course credits. All participants reported normal colour vision, and normal or corrected-to-normal visual acuity. No participants were excluded prior to the analysis.

#### *5.6.1.2. Materials and apparatus*

Six agent images from the previous experiments were used in Experiment 6, three female images and three male images (see Figure 5.7). Each agent was associated with a unique voice. The voice recordings of instructions were delivered by native English speakers. We labelled each agent with a number in order to explain the experimental design more clearly.

Each object can be identified by two distinct features: colour and shape. Nine colours and thirteen shapes were used to generate a pool of totally 117 different objects. The colours used in the current study were: red, orange, yellow, green, blue, purple, grey, brown, and pink. The shapes were: bag, cap, clock, mug, glasses, notebook, trousers, scarf, t-shirt, shoes, socks, tie, and umbrella. Examples of objects are shown in Figure 5.8.

The experiment was presented on a 14-in Samsung laptop. E-prime 2.1 (Psychology Software Tools, Pittsburgh, PA) was used for the presentation of all materials and for recording response time and accuracy.


Figure 5.7. The agent images used in Experiment 6.

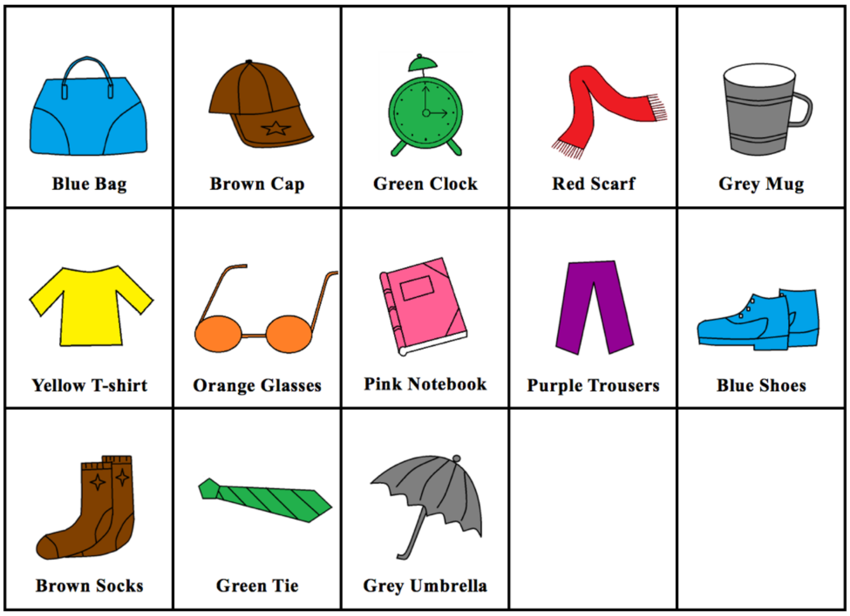


Figure 5.8. Examples of the object images used in Experiment 6.

#### *5.6.1.3. Design and Procedure*

Participants were first presented with a video to get familiarised with their communicative partners, i.e., the agents. The video lasted for 95 seconds, depicting each agent coming in to a room through a wooden door, dancing for approximately 10 seconds and then leaving the room through the door. Figure 5.9 illustrates some of the frames extracted from the original video.

Figure 5.9. Frames extracted from the original video used in Experiment 6.

A total number of 36 grid images were presented in the main experiment, thus resulting in 36 experimental trials. These 36 experimental trials were grouped into 6 blocks each containing 6 trials. Participants practiced on 4 additional trials prior to the main experiment. Every experimental trial consisted of two phases: the presentation phase and test phase.

In the presentation phase, a 4x4 grid was presented to participants with six objects placed on it (see Figure 5.10). Every object on the grid had its unique colour and shape. These six objects were further divided into two sets: set A and set B, each containing three objects. The first agent came into the room and stand behind the grid. He or she then gave three individual instructions which referred to the three objects in set A. The instructions were structured as ‘click on the + colour + shape’, e.g. ‘click on the pink mug’. Participants were asked to follow the agents’ instructions and click on the referred objects on the grid using a computer mouse. Participants were given 4000ms to make a response to each instruction from the onset of the instruction. If no response was made within 4000ms, then a time-out would trigger the start of next instruction or next presentation. After giving three instructions, the first agent went out of the room through the door. Then either the same agent (i.e., *the no-change condition*) or a new agent (i.e., *the change condition*) came into the room and gave another three instructions which referred to the three objects in set B. After all the instructions were given, the six objects on the grid were blocked off by 6 black squares, which marked the end of the presentation phase.

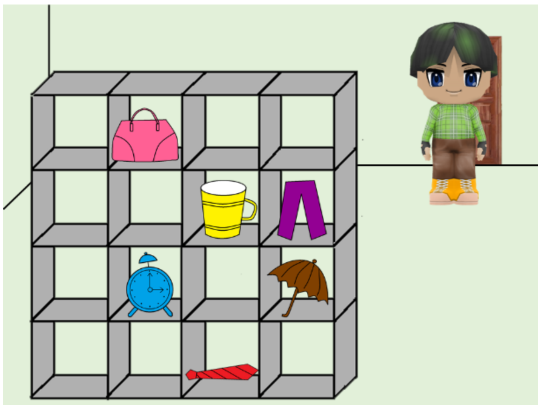

Figure 5.10. An example of the grid image during the presentation phase.

After the six objects were occluded by black squares, a test phase began with six single-recognition-probe tasks (see Figure 5.11). Six probe images were presented sequentially on a white test board in the lower right corner of the screen. The probes contained both target probes and foils. Participants were instructed to response 'yes' to target probes which were objects previous displayed on the grid on the same trial. Participants were instructed to response 'no' to foil probes which were not objects presented on the grid on the same trial. Foil probe was always made of new colour and shape which had not been presented in the same experimental trial either as objects displayed on the grid or as probes.

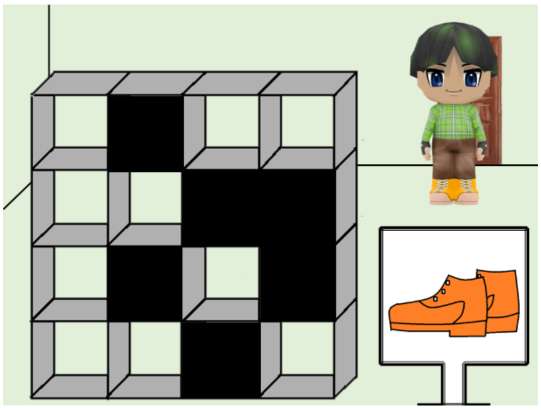

Figure 5.11. An example of the single-recognition probe task presented during the test phase.

Within each block, there were always three trials in the change condition and three trials in the no-change condition. Participants never encountered more than two trials of the same condition successively. Objects in Set A were always introduced by Agents 1, 2, or 3, and each of the three agents was equally likely to appear. In the no-change condition, objects in set B were introduced by the same agents as set A. In the change condition, objects in set B were introduced by Agent 4, 5, or 6, and each of these three agents was equally likely to appear. No agent appeared in two successive trials. All possible pairings between Agent 1/2/3 and Agent 4/5/6 in the change condition were presented, and each pair of agents occurred twice during the whole experiment.

Target probes could be either objects from set A or set B. The number of foils, the number of target probes from set A, and the number of target probes from set B, each ranged between 1 to 3 objects. This resulted in 6 possible compositions of the probes (see Table 5.3). This was designed to prevent participants from predicting answers and being strategic when performing the task. These six possible combinations were randomly assigned to the six experimental trials within each block. Objects introduced at various points of the presentation phase (first, second, or third in each set) were equally likely to be selected as target probes and were equally likely to be presented at all probe positions (from the first to the sixth). In the whole experiment, each object appeared no more than twice during the presentation stage, no more than twice as a target probe and no more than once as a foil during the test stage. No object was presented twice on any three successive trials.

Table 5.3. The possible compositions of the six probes on each experimental trial.

| Number of Objects from Set A | Number of objects from Set B | Number of Foils |
|---|---|---|
| 1 | 2 | 3 |
| 2 | 1 | 3 |
| 1 | 3 | 2 |
| 3 | 1 | 2 |
| 2 | 3 | 1 |
| 3 | 2 | 1 |

### 5.6.2. Results

As we aimed to investigate the effect of agent change on participants' memory for objects they recently interacted with, our main interest was to examine whether the change of agent identity in change condition would have an effect on participants' memory for the objects from set A. The reason was that only objects from set A were probed either in the presence of the same agent who verbally mentioned the object or in the presence of a new agent, whereas objects from set B were always probed in the presence of the same agent. Therefore, only the accuracy and response time data collected for set A probes were entered into the main analysis. Participants' overall accuracy was used to check against chance level, in order to make sure that all participants included in analysis understood our experimental instructions, were not making random responses throughout the experiment and did not have major memory related problems. All participants performed above chance, thus no participant was excluded prior to analysis. For the response time data analysis, only the trials to which participants responded correctly were included. Then response times higher than 2.5 standard deviations from mean response time were excluded, which caused 1.7% of the response time data for correctly responded trials to be dropped. No response was found to be quicker than 200ms.

Paired t-tests were carried out between the no-change condition and change condition on both accuracy and response time data from set A. For accuracy, no difference was found between the no-change condition (*mean* = .79, *SD* = .12) and the change condition (*mean* = .77, *SD* = .12), $t(29) = 1.275$, $p = .212$. Similarly, for response time, there was no difference between the no-change condition (*mean* = 688ms, *SD* = 120) and the change condition (*mean* = 692ms, *SD* = 113), $t(29) = -.509$, $p = .614$. Overall, the results showed no effect of the change of agents on participants' memory performance.

### 5.6.3. Discussion

In the present experiment, participants performed virtually identically in the change and no-change conditions. This suggests that the change of conversational partner during the presentation phase did not disrupt participants' memory for objects in the memory test. No evidence was found to support the hypothesis that the change of conversational partner could structure memory in ways similar to the effect of spatial shift on memory in Radvansky and his colleagues' studies (Radvansky & Copeland, 2006; Radvansky et al., 2011, 2010).

However, it is noteworthy that although only data from set A were entered into analysis, participants were not aware of this analysis plan. Participants were instructed that they would be examined on whether they could correctly recognise the true probes from both set A and set B and correctly reject the foils, which required them to encode all the six objects on each trial. Therefore, for the change condition trials, having a change of conversational partners in the middle of participants' encoding phase might have helped participants' memory for objects, rather than disrupted their memory as we previously predicted (e.g., Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016; Radvansky, 2012). According to Radvansky (2012) and Pettijohn et al. (2016), the spatial shift of 'walking through doorway' can cause not only forgetting, but remembering. Specifically, Radvansky points out that the presence of an event boundary during encoding improves people's overall memory, because it decreases retroactive interference (which is a memory phenomenon that the newly learned information interferes with the recall of previous learned information). For example, Experiment 1 in Pettijohn et al. (2016) asked participants to learn two lists of 20 words, one list in the shift condition and the other in the no-shift condition. They first learned the first 10 words of one list in one room, and then either moved to another room (shift condition) or across a large room (no-shift condition) where they learned the last 10 words of that list. A memory test was presented two minutes after either a spatial shift or no shift, and participants were asked to recall the entire lists. Results showed that participants' memory for the whole word list was improved when they experienced a spatial shift in the middle of the learning phase compared to when they simply moved across a large room. This finding suggests that if participants experience an event shift in the middle of their encoding, their memory of the encoded information might be enhanced. Therefore, we decided to run a follow-up experiment in which the event boundary would be moved from the middle of encoding phase to the point after encoding has been completed, and to examine whether the predicted effect of the change of conversational partners on memory could then be observed.

## 5.7. Experiment 7

### 5.7.1. Methods

#### *5.7.1.1. Participants*

Twenty-one students (17 females, mean age 19.04, age range 18 to 29) from the University of Birmingham participated in this study in return for course credits. All participants reported normal colour vision, and normal or corrected-to-normal visual acuity. No participants were excluded prior to the analysis.

#### *5.7.1.2. Materials and apparatus*

The materials and apparatus used in Experiment 7 were identical to those of Experiment 6 with the only exception that the objects in Experiment 7 contained twelve different shapes rather than thirteen shapes. Nine colours (i.e., red, orange, yellow, green, blue, purple, grey, brown, and pink) and twelve shapes (bag, cap, clock, mug, glasses, notebook, trousers, scarf, t-shirt, shoes, socks, and tie)[9] were used to generate a pool of 108 different objects. Each object appeared only 2 or 3 times during the presentation stage.

#### *5.7.1.3. Design and Procedure*

The design and procedure were identical to that of Experiment 6 with the following exceptions. In the presentation phase, all six objects from set A and set B were introduced by the same agent. After giving six verbal instructions, the agent went out of the room. Then either the same agent (i.e., *the no-change condition*) or a new agent (i.e., *the change condition*) came into the room, followed by the memory test.

### 5.7.2. Results

In the current experimental design, the change of agent happened after the first agent introduced all six objects. The data for true probes and foil probes were processed separately, as they involved different types of responses. For the response time analysis, only the trials to which participants responded correctly were included. Response times longer than 2.5 standard deviations from mean response time were excluded prior to the analysis, causing 3% of the response time data from correctly responded trials to be dropped. No response was found to be quicker than 200ms.

---

[9] Umbrella was removed from the pool of objects due to the caution that the difficulty level of remembering umbrellas might be higher than that of remembering other objects, as indicated by some participants' feedback during the debriefing session and also by the by-item analysis we did on participants' accuracy and response time data in Experiment 6.

The descriptive data of accuracy and response time are summarised in Table 5.4. For true probes, paired t-tests were carried out between the no-change condition and change condition on accuracy and response time data respectively. For accuracy, there was no difference between the change condition and the no-change condition, $t(20) = 1.008$, $p = .325$. For response time, no difference was found between the change condition and the no-change condition $t(20) = .400$, $p = .693$. For foils, similarly, paired t-tests were carried out between the no-change condition and change condition on accuracy and response time data respectively. For accuracy, participants made less errors in the change condition than in the no-change condition, $t(20) = 2.137$, $p = .045$. For response time, a marginally significant difference was found between the change condition and the no-change condition, $t(20) = 1.886$, $p = .074$, and this non-significant effect suggested that participants tended to take longer time to make a response in the no-change condition compared the change condition.

Table 5.4. Accuracy (proportions) and Response Time (ms) with standard deviations for Experiment 7.

| | **Accuracy** | | | | **Response Time** | | | |
|---|---|---|---|---|---|---|---|---|
| | No-change | | Change | | No-change | | Change | |
| | M | SD | M | SD | M | SD | M | SD |
| **True probes** | 0.83 | 0.09 | 0.84 | 0.10 | 738 | 108 | 743 | 124 |
| **Foils** | 0.90 | 0.08 | 0.93 | 0.09 | 782 | 137 | 768 | 126 |

### 5.7.3. Discussion

Experiment 7 was conducted to examine whether the predicted effect of the change of conversational partners on memory could be observed after controlling for a potential confounding effect of having an event abound during encoding. Findings from Experiment 7 showed that participants performed similarly in the change condition and the no-change condition when it came to true probes. Participants even performed better in the change condition compared to the no-change condition when it came to foil probes (which was opposite to our hypothesis). No evidence was found to support our hypothesis that change of agent would hinder people's memory performance on objects appeared before the agent change. Detailed discussion on the implications of these findings will follow in the next section.

## 5.8. General Discussion

### 5.8.1. Summaries of Experiments 4-7

Our study in Chapter 5 was the first to integrate the theory of the low-level memory mechanism for common ground representations (Horton & Gerrig, 2005a, 2005b) and the theory of the memory processes involved in event model representations (e.g., Radvansky & Zacks, 2014) to propose a plausible route of how partner-specificity common ground can be achieved. Two new experimental paradigms were designed to seek evidence for the proposed low-level memory mechanism for partner-specific representations, which may be shaped by the relatively automatic event segmentation system. An innovative computerised 'shopping task' was employed in experiments 4 and 5, and an adapted version of the director task was employed in experiments 6 and 7. In both experimental paradigms, the key manipulation was whether the memory probes appeared during the presence or the absence of the specific conversational partner who were associated with the probed objects. Our key investigation was whether the change of conversational partners prompted an event boundary which consequentially made recently encoded objects less accessible in memory, compared to when such change did not happen. In our study, this change of conversational partners was irrelevant to the task participants were performing, and participants were never explicated asked to pay attention to the identity of the conversational partner. If an effect of partner change could be observed even when such an effect would be detrimental to participants' task performance, then it gains plausibility for the view that partner-specific representations can be encoded, tracked and accessed in a relatively automatic way.

The current findings from Experiments 4 to 7 together, however, do not support our initial hypothesis. Experiment 4 provides preliminary results in support of our hypothesis that people's memory for objects become less available after experiencing a change of agents compared when such change has not happened. Experiment 5 failed to replicate the same effect after controlling for the confounding factor of visual distraction introduced by agent-facing scenarios. Some convergent evidence seemed to be gained from participants' responses on trials with current object probes and foils, but responses on trials with previous object probes revealed a pattern of results that contradicted our hypothesis. Experiment 6, where the previous object probes were removed, failed to reveal any agent-change effect on participants' memory. Experiment 7 further controlled for a cofounding factor, i.e., the occurrence of an event boundary in the middle of the whole object encoding phase. However, Experiment 7 again failed to observe any effect in line with our hypothesis. Taken together, the findings from Experiments 4 to 7 indicated a minimal and inconsistent role of the change of agent identity on participants' memory under the current experimental settings.

### 5.8.2. Formation and retrieval of partner-specific associations: automatic?

Participants do form associations between conversational content and relevant conversational partners and represent these partner-specific associations in episodic memory in some circumstances, as they exhibit partner-specific effects in certain communicative situations (Brennan & Clark, 1996; Brown-Schmidt, 2009a; Horton, 2007). The key interest of the current study lies in the processes of how the partner-specific representations are formed, tracked and accessed in memory. We hypothesised that people may form partner-specific associations by automatically segmenting events into partner-labelled episodic chunks, which may in result implicitly affect the availability of objects, even when it is unnecessary and detrimental to do so. However, the findings from Experiments 4 to 7 did not support this hypothesis. Some people might argue that the current experimental paradigms are not sensitive enough to detect the predicted effect of agent-change on memory; however, I argue that this insensitivity itself might be informative in telling us something about the encoding and retrieval mechanism for partner-specific representations.

As detailed in sections 4.2 and 4.3, the low-level memory-based mechanism for common ground proposed by Horton & Gerrig (2005a, 2005b) primarily focuses on a cue-based memory retrieval process upon which common ground can be accessed relatively automatically, quickly and effortlessly. Many studies have provided evidence in support of such a low-level memory retrieval process of common ground (Barr, 2014; Gorman et al., 2013; Horton, 2007; Horton & Slaten, 2012). For example, Horton (2007) demonstrated that speakers was faster to name an object when they were in the presence of an individual with whom the object name was previously associated. However, a recent study by Brown-Schmidt and Horton (2014) failed to replicate the key findings of the first of the two experiments in Horton (2007). This failure of replication casts doubt on the idea that the mere associations between partners and referents would be sufficient to facilitate participants' access to the object names in memory. Horton and Gerrig (2016) also revisited the memory-based processing approach to common ground recently. Horton and Gerrig (2016) acknowledged in the paper that the partner-referent associations in Horton (2007) was relatively arbitrary as the only link between a partner and a referent was that participants named an object while the partner was seated next to the participant without any interaction. This weak link might have been insufficient for partner identity to act as a very salient cue to active relevant memory representations.

Moreover, the low-level memory mechanism for common ground (Horton & Gerrig, 2005a, 2005b) is primarily proposed on the basis of assuming that associations between conversational partners and referents have been formed and stored in memory. However, Horton and Gerrig did not specify the memory encoding process of how the partner-specific associations can be formed in memory: for example, is the formation of partner-specific associations automatic, fast and effortless, or is it constrained by cognitive resources and cognitive goals? A recent paper by Brown-Schmidt, Yoon, and Ryskin (2015) emphasises on the encoding process of partner-specific associations, and proposes that people form partner-specific representations under the constraints of attention, memory and communicative relevance. Brown-Schmidt et al. (2015) agree with the idea that conversational partners act as a special type of context and thus provide contextual memory cues for conversational participants to retrieve what is shared between them (Horton & Gerrig, 2005a, 2005b), but they believe that only 'a limited number of bindings' should be encoded between the specific information and its contextual cues. This proposal is against an automatic encoding mechanism for partner-specific association for two reasons. Firstly, Brown-Schmidt et al. (2015) argue in the paper that if people automatically form all contextual associations as 'an unbounded number of bindings' will be formed which exceeds the limits of attention and memory. Secondly, they propose that partner-specific representations should be constrained by the communicative relevance of the conversational partner. Namely, when a partner is communicatively relevant, or when there is extensive partner interaction, the bindings between the content of the conversation and the partner would be formed more strongly and tightly in memory.

Drawing upon this up-to-date literature on the memory-based mechanism for common ground, there are at least two ways in which we could explain the lack of partner-specific effect in our current study. The first plausible explanation is that participants did not form partner-specific associations in the mere co-presence of the conversational content and conversational partner. One reason is that the identity of the agents was not highly communicatively relevant in our experimental paradigms, and participants might have been busy remembering /rehearsing the objects rather than paying attention to the agent identities. The partner-specific representations may not be formed in a completely automatic, stimuli-driven way, although they can still be formed spontaneously and efficiently when the conversational partner serves as a meaningful and relevant conversational context. Another plausible explanation is that although participants have formed associations between specific partners and objects, these associations might not be strong and sufficient to facilitate participants' memory retrieval of the objects. What is in common between these two explanations is that they both include the notion that the memory-based processes of partner-specific representations are constrained by the communicative relevance/meaningfulness of a partner in a conversational context. The difference is that the first explanation focuses on the formation of partner-specific representations, while the second one focuses on the retrieval of partner-specific representations. This emphasis on the communicative relevance of the conversational partner might also may explain the reason why Radvansky and colleagues' studies (Radvansky & Copeland, 2006; Radvansky et al., 2011, 2010) have managed to observe an effect of spatial shift, but our study failed to observe an effect of agent-change. In Radvansky et al.'s study, participants were actively experiencing the spatial shift, as they were performing a walking action to cross the space in a virtual environment. In our study, however, the change of agents was only presented on a screen in front of participants and thus were only passively experienced by participants. Future research is needed to further examine these two explanations and to shed light on the specific processes of the proposed memory-based mechanism for common ground.

### 5.8.3. Limitations and future improvements

Firstly, as discussed in the previous section, the change of agents in the current series of experiments was only presented on a screen in front of participants and thus were only passively experienced by participants. Future research could try to adapt the experimental paradigm to make the conversational partners more communicatively relevant and meaningful, in order enhance participants' communicative motivation. For example, we could adapt the director task used in Chapters 2 and 3 for which participants had to use their conversational partner's perspective to interpret their messages by implementing the change of directors into the experimental paradigm.

Secondly, it was suggested by some researchers that the memory tests in the present series of experiments might be very difficult due to the fact that many of the presented objects were from the same semantic category, i.e., clothing. The high level of difficulty of the memory tests might have caused the current experimental paradigms to be insensitive to the predicted effect of agent-change on memory. However, participants' overall accuracy was in fact relatively high in Experiments 4 to 7 (i.e., the mean accuracy for each probe type in each experiment was all above 0.77), which demonstrated that it was unlikely that participants found the memory tests very difficult due to objects being from the same semantic category. Nonetheless, future research could be done to use stimuli from different semantic categories to control for this confounding variable.

Thirdly, the memory probes occurred repeatedly during the current experiments, and participants were highly likely to have been prepared for the memory probes and have used memory strategies like rehearsal during the encoding phase (although they could not anticipate exactly at which point the memory test would occur in Experiment 4 and 5). Additionally, the duration between the encoding phase and the retrieval phase was relatively short, so that the memory tests were likely to have tapped into participants' working memory rather than longer term memory. Future research could consider using an unexpected recognition task to test participants' memory for objects after they have completed the whole communication task. In this case, participants are not aware of this memory test when they are undertaking the preceding communication task. They are therefore unlikely to have employed any memory strategy in order to enhance their memories for the objects occurred in the conversation. The duration between the encoding phase and the retrieval phase would be relatively long, which would make the memory test more likely to tap into participants' long-term memory rather than working memory.

## 5.9. Conclusion

We propose that people may automatically segment events into partner-indexed episodic memorial chunks through which partner-specific associations are automatically encoded in memory, and these partner-specific associations may affect the availability of objects associated with the partners. However, Experiments 4-7 provided little support for this hypothesis. The findings cast doubt on the idea of a completely automatic memory-based mechanism for common ground, and provide incentive for future research to further examine the potential constraints on the formation and retrieval process of partner-specific representations in language processing.

# CHAPTER VI

# Is memory for objects influenced by their being encoded from common ground versus privileged ground?

## 6.1. Introduction

Researchers have long assumed that common ground is the foundation upon which people communicate with each other (e.g., Clark & Marshall, 1978, 1981). It is, therefore, important and interesting to explore how people spontaneously assign attention (and possibly other cognitive resources) to information in common ground versus privileged ground during communication, and how this affects their memories for information in common ground versus privileged ground in a post-test. The aim of Chapter 6 was to investigate whether the long assumed privileged status of common ground over privileged ground could be achieved via a relatively low-level memory-based mechanism. Specifically, we were interested in whether participants' memory for objects in a later unexpected memory test would be affected by the mere difference of whether objects occurred in the common ground or privileged ground during a preceding communication task. Our rationale was: if more attention was automatically assigned to objects in the common ground compared to objects in the privileged ground when participants were engaged in the communication task, then we would expect participants to show higher recognition rates for objects in the common ground than those in the privileged ground in the following memory test. If participants automatically showed higher recognition rates for objects in the common ground compared to objects in the privileged ground, then this effect would provide some evidence in favour of the proposal that common ground information might be represented and assessed via a relatively low-level memory mechanism.

The present study also examined whether participants' memory for objects would be affected by how the common ground was established. As introduced in Chapter 1, Clark and Marshall (1981) have proposed that common ground can be established on three bases, i.e., *physical co-presence, linguistic co-presence* and *cultural co-presence/community membership*. Information that either exists in the shared physical environment (physical co-presence), or is maintained in the shared past or present conversations (linguistic co-presence), or is part of the shared sociocultural background between speakers and listeners (community membership) can be regarded as mutual knowledge between conversational partners. In the present study, we used both physical co-presence and linguistic co-presence to establish common ground shared between participants and their conversational partner in the communication task. As a result, each object could bear one of three types of ground status: common ground established by both physical and linguistic co-presence, common ground established merely by physical co-presence, and privileged ground. For example in Figure 6.1, the rabbit was visible to the director and was also referred to by her, so it could be considered as in the type of common ground established by both physical and linguistic co-presence. The sunglasses were visible to the director but were not referred to by her, therefore in the common ground established by merely physical co-presence. In contrast, the pineapple was invisible to the director therefore in participants' privileged ground.

A particular interest of the present study lied in whether participants' memory for objects in the common ground established by mere physical co-presence would be more accurate than their memory for objects in the privileged ground. We expected participants' memory for an object to be facilitated merely by its being presented in the common ground than in the privileged ground. We also expected participants to show more accurate memory for object in the common ground established by both physical and linguistic co-presence compared to objects in the common ground established by mere physical co-presence, as the linguistic interaction might add details and richness to the representations of the common ground information.

Figure 6.1. An example of the shelf images used in this study.

# 6.2. Experiment 8

## 6.2.1. Methods

### *6.2.1.1. Participants*

Thirty-two students (24 female, mean age 21.63, age range 17 to 32) from the University of Birmingham participated in this study in return for study credits. All participants had normal colour vision and normal or corrected-to-normal visual and auditory acuity.

### *6.2.1.2. Design and procedure*

During the experiment, participants first played an adapted version of the director task. Participants then took an unexpected memory test in which they were asked to recognise whether the object images presented in the memory test had been seen in the director task. Both the director task and the memory test were presented with E-prime 2 software (Psychology Software Tools, Pittsburgh, PA) on a lab computer.

**Director task**

Participants were first invited to play a variant of the director task (which was similar to the visible perspective version of the director task in Chapter 3). During the instructions phase, participants were made aware of the perspective difference between a director and themselves, and were instructed that if the director could not see a certain object, then she could not ask participants to move the object. Participants were instructed to bear the director's perspective in mind when interpreting her instructions, and were given an example of how the director's limited perspective could constrain her reference. Each trial began with a fixation stimulus (+) displayed in the centre of the screen for 1000s. After the 1000ms fixation phase, an image of a 4 x 4 shelf was presented. A female director was standing behind the shelf. There were always 12 objects on the shelf, one object in a slot. All objects were items we encounter in our everyday life, from categories such as food, animal, cookware, clothing, instruments and toys. Of the 12 objects on the shelf, there were always 8 objects visible to both the participants and the female director, and thus in their common ground. The remaining 4 objects were occluded from the director's perspective by green squared backgrounds, and thus in participants' privileged ground. Participants were first given 5000ms to view the shelf image, following which they were given instructions by the director to move objects on the shelf. Participants were instructed to respond to the instructions as quickly and accurately as possible. A new instruction was always presented 8000ms after the onset of the previous instruction in order to control for encoding time, regardless of whether a response was received.

Sixteen shelf images were presented to participants sequentially during the director task, each shelf image was accompanied by 4 instructions and each instruction referred to an unique object on the shelf. The 16 shelf images formed 2 test blocks, each block with 8 shelf images. Participants were allowed to take a break between the blocks. Two different versions of the director task were created by reversing the order of the presentation of the 16 shelf images, i.e., an original order version and a reversed order version. Before the experimental phase, participants were presented with 2 additional shelf images for practice, each shelf image consisted of 2 instructions. Each object image was only presented once across the entire task.

Among the 16 shelf images used in the experimental phase, 8 shelf images each consisted of a set of critical objects which were 3 objects of the same exemplar but of different sizes (e.g., a small balloon, a medium balloon and a large balloon) and 9 objects from different exemplars (see Figure 6.2a). Each of these 8 shelf images contained 1 critical instruction and 3 non-critical instructions. A critical instruction (e.g., 'move the big balloon one slot down from there') required participants to use the director's perspective to identify the correct referent from the 3 objects of the same exemplar (as the biggest balloon among the three balloons were occluded from the director, and thus could not be referred to by the director). The position of the critical instruction varied from the 1$^{st}$ to the 4$^{th}$ of a set of instructions presented with each shelf image. For the other 3 non-critical instructions, participants did not have to use the director's perspective to resolve the reference. The other 8 shelf images each consisted of 12 objects from different exemplars (see Figure 6.2b), and 4 non-critical instructions were given with each shelf image. The structure of the critical instructions was always 'move the [scalar adjective] [noun] one slot [directional word] from there' (e.g., nudge the small ball one slot up); whereas the structure of non-critical instructions was either 'move the [scalar/normal adjective] [noun] one slot [directional word] from there' or 'move the [noun] one slot [directional word] from there'. Directional words were 'left' (17 out of 64 instructions) 'right' (17 instructions) 'up' (15 instructions) and 'down' (15 instructions). Participants were told that the directional words 'left' and 'right' referred to their own left and right sides. For critical instructions, only ‘up’ and ‘down’ directional words were used.

Figure 6.2. (a) The left shelf image contained 3 objects of the same exemplar (i.e., 3 balloons of different sizes) and 9 objects from different exemplars. (b) The right shelf image contained 12 unique objects from different exemplars.

**Memory test**

After the director task, participants were asked to undertake an unexpected memory test in which they were asked to recognise the objects that had been presented on the shelf. None of the three objects of the same exemplar was included in the memory test. The memory test consisted of 48 true probes and 48 foil probes. The 96 probes were randomly selected to be presented sequentially to each participant. The 48 true probes consisted of 16 objects which were in the common ground and had been referred to by the director (which was labelled as *linguistic common ground objects*, one from each shelf image), 16 objects which were in the common ground but had not been referred to by the director (labelled as *visual common ground objects*, one from each shelf image), and 16 objects in the privileged ground (labelled as *privileged ground objects*, one from each shelf image). The 48 foil probes were also chosen from everyday items from categories such as food, animal, cookware, clothing, instruments and toys. Each probe trail began with a fixation stimulus (+) displayed in the centre of the screen for 1000 ms. Following the fixation screen, participants were presented with a probe image, to which they were instructed to make a response by either pressing the right arrow key labelled as ‘yes’ if they had seen the object during the director task, or pressing the left arrow key labelled as ‘no’ if they had not seen the object during the director task. The next trial only commenced after a response had been made. Participants were instructed to respond as accurately as possible. Although both participants’ response time and accuracy were recorded during the memory test, our main focus was on accuracy. Therefore, participants were not required to respond as quickly as possible.

### 6.2.2. Results

**Director task**

The director task was mainly used as an encoding phase in the present study. Our focus was not on participants’ performance in the director task but on their performance in the memory test. As a result, only the accuracy rates of participants’ responses to the critical instructions of the director task were examined as a performance check. Among the 32 participants, 1 participant committed 7 egocentric errors out from the 8 critical instructions. This high rate of egocentrism was likely to be due to a misunderstanding of the task or a lack of concentration, so this participant’s data from the memory test were excluded prior to our analyses. Other 31 participants all performed relatively well in the director task (average accuracy was 0.984), with 4 participants making 1 egocentric error each and others making no egocentric errors.

**Memory test**

Participants whose overall accuracy in the memory test failed to be significantly above chance level (i.e., failed to correctly response to 57 or more probes out of 96 probes) would be excluded from the analyses on both accuracy and response time. In the present experiment, all participants' overall accuracy was significantly above chance level, ranging from 0.625 to 1, so no one was excluded prior to the following analysis.

The means and standard errors of the accuracy data for each probe type were reported in Tabl3 6.1. A one-way repeated-measures ANOVA was conducted on accuracy with probe type (i.e., linguistic common ground object, visual common ground object, and privileged ground object) as a within-subject factor. A main effect of probe type was observed, $F(2, 60) = 74.365$, $p < .001$, $\eta p^2 = .713$. Two planned paired t-tests were carried out between recall rates of linguistic common ground objects and those of visual common ground objects, and between recall rates of visual common ground objects and those of privileged ground objects, to examine if there were incremental differences among these three probe types as we predicted. Paired t-tests revealed that participants' recognition memory for linguistic common ground objects was significantly more accurate than that for visual common ground objects, $t(30) = 9.401$, $p < .001$, *Cohen's d* = 1.648. Participants' recognition memory for visual common ground objects was significantly more accurate than for privileged ground objects, $t(30) = 2.381$, $p = .024$, *Cohen's d* = 0.446. The significance level for each t-test was adjusted to .025 after applying Bonferroni corrections.

For the response time analysis, only the trials to which participants responded correctly were included. Response times longer than 2 standard deviations from mean response time were excluded prior to the analysis, which caused 3.2% of the response time data from correctly responded trials to be dropped. No response was found to be quicker than 200ms. The means and standard errors of the response time data used in the analyses for each probe type were reported in Table 6.2. Similar to the accuracy data, one-way repeated-measure ANOVA was carried out on response time with probe type as a within-subject factor. No main effect of probe type was found, $F(2, 60) = 2.323$, $p = .107$, $\eta p^2 = 0.072$.

Table 6.1. Means and standard errors of participants' accuracy (in proportions) for each type of true probes and foils probes, for Experiment 8, 9 and 10 respectively.

| | True probes | | | | | | Foil probes | |
|---|---|---|---|---|---|---|---|---|
| | Linguistic common ground objects | | Visual common ground objects | | Privileged ground objects | | | |
| | Mean | SE | Mean | SE | Mean | SE | Mean | SE |
| Experiment 8 | 0.821 | 0.024 | 0.536 | 0.037 | 0.452 | 0.031 | 0.929 | 0.011 |
| Experiment 9 | n/a | n/a | 0.585 | 0.028 | 0.550 | 0.026 | 0.949 | 0.016 |
| Experiment 10 | n/a | n/a | 0.580 | 0.025 | 0.600 | 0.022 | 0.947 | 0.009 |

Table 6.2. Means and standard errors of participants' response times (in milliseconds) for each type of true probes and foils probes, for Experiment 8, 9 and 10 respectively.

| | True probes | | | | | | Foil probes | |
|---|---|---|---|---|---|---|---|---|
| | Linguistic common ground objects | | Visual common ground objects | | Privileged ground objects | | | |
| | Mean | SE | Mean | SE | Mean | SE | Mean | SE |
| Experiment 8 | 1058 | 56 | 1119 | 81 | 1162 | 76 | 1124 | 70 |
| Experiment 9 | n/a | n/a | 976 | 42 | 961 | 38 | 1053 | 57 |
| Experiment 10 | n/a | n/a | 922 | 34 | 898 | 30 | 936 | 42 |

### 6.2.3. Discussion

The current findings confirmed our main hypothesis that participants would show significantly better memory for linguistic common ground objects compared to visual common ground objects, and also significantly better memory for visual common ground objects over privileged ground objects. These findings provide some evidence to the idea that people may automatically form stronger memory representations for common ground information over privileged ground information during communication.

One limitation of the current findings was that the difference observed between the recognition rates of visual common ground objects and privileged ground objects, though significant ( $p = .024$), was not strong, as its effect size was only modest (*Cohen's d* = .446). As the main focus of the present study was to examine the difference between participants' memories for visual common ground objects and for privileged ground objects, we decided to collect more data on this pair of probe types in a follow-up experiment. If the significant difference between participants' memories for visual common ground objects and privileged ground objects could be replicated in the follow-up experiment, then we would be more confident to pursue further research ideas upon this reliable effect.

Additionally, an alternative explanation was that the significant difference between visual common ground objects and privileged ground objects might merely be driven by an item effect rather than a genuine effect of the objects' ground status. Since we did not use the very same objects for visual common ground and privileged ground, it was possible that the visual common ground objects were overall easier to be remembered and retrieved by nature compared to the privileged ground objects, which contributed to the effect observed in Experiment 8. In order to examine this alternative explanation regarding item effect, we decided to create another version of the director task alongside the use of the original version in the follow-up experiment. The visual common ground objects from the original version became privileged ground objects in the new version, and the privileged ground objects from the original version became visual common ground objects in the new version. Participants took part in either the original version or the swapped version of the director task. If the follow-up experiment replicated the significant effect between these two different probes types in both task versions, then this alternative explanation could be dismissed.

## 6.3. Experiment 9

### 6.3.1. Methods

#### *6.3.1.1. Participants*

Twenty-six students (16 female, mean age 21.69, age range 19 to 39) from the University of Birmingham participated in this study in return for a small honorarium or study credits. Fourteen participants took the original shelf version of the director task and correspondingly the stimulus set A of the memory task, twelve participants took the swapped shelf version of the director task and the stimulus set B of the memory task (see section 6.3.1.2 for more details). All participants had normal colour vision and normal or corrected-to-normal visual and auditory acuity.

#### *6.3.1.2. Design and procedure*

The design and procedure were identical to that of Experiment 8 with the following exceptions. A new version of the director task was created by swapping the ground status of the visual common ground objects and the privileged ground objects presented on each shelf image in Experiment 8 (see Figure 6.3). For the 8 shelf images without critical instructions, the ground status of all four visual common ground objects and four privileged ground objects was swapped. For the other 8 shelf images with critical instructions, the ground status of a pair of objects (one is privileged ground, and one in visual common ground) remaining unchanged for each shelf, as the critical object in the privileged ground needed to remain in the privileged ground for the critical instruction to work appropriately. For example, the red small ball in the privileged ground in Figure 6.3a remained unchanged in the privileged ground in Figure 6.3b. Both the original shelf version and the swapped shelf version were used in Experiment 9. Each version generated two sub-versions by reversing the order of the presentation of the shelf images. Participants were assigned to one of the four task versions:

original-shelf-original-order, original-shelf-reversed-order, swapped-shelf-original-order, and swapped-shelf-reversed-order.

Figure 6.3. (a) example of the original shelf image used in Experiment 8. (b) Example of the new shelf image creased for Experiment 9 by swapping the visual common ground objects and the privileged ground objects presented on the shelf.

The memory test consisted of 112 true probes and 48 foil probes. The true probes contained 56 visual common ground objects and 56 privileged ground objects. We did not include the linguistic common ground objects in the true probes, as our main interest was whether participants showed a difference between their memory for the visual common ground objects and memory for the privileged ground objects. All the visual common ground object and privileged ground objects presented on the shelves were included in the memory test, except for the un-swapped pairs of objects from the shelf images involving critical instructions. The 48 foils remained as foils in Experiment 8. Two different 'stimulus sets' (i.e., stimulus set A & stimulus set B) were created for the memory test to compliment the different versions of the director task. The visual common ground objects in stimulus set A were the privileged ground objects in stimulus set B, and vice versa. For participants who took the original-shelf-original-order or original-shelf-reversed-order versions of the director task, they later participated in the memory test with stimulus set A. For participants who took the swapped-shelf-original-order or swapped-shelf-reversed-order versions of the director task, they later participated in the memory test with stimulus set B.

### 6.3.2. Results

**Director task**

Among the 26 participants, 4 participants committed 1 egocentric error each out of the 8 critical instructions, and the other participants did not commit any egocentric errors. The overall accuracy rate is .995.

**Memory test**

Three participants whose overall accuracy in the memory test failed to be significantly above chance level (i.e., failed to correctly respond to 91 or more probes out of 160 probes) were excluded prior to the following analyses.

Participants' performances on the memory test with stimulus set A (11 participants) and stimulus set B (12 participants) were analysed together, as this factor was not the interest of the present study[10]. The means and standard errors of the accuracy data for each probe type were reported in Table 6.1. Paired t-test was conducted between participants' accuracy of visual common ground objects and privileged ground objects. Although participants' recognition memory for visual common ground objects showed a tendency to be better than their recognition memory for privileged ground objects, no significant difference was observed, $t(22) = 1.396$, $p = .177$, *Cohen's d* = .265.

For the response time analysis, only the trials to which participants responded correctly were included. Response times longer than 2 standard deviations from mean response time

---

[10] When a mixed 2 x 2 ANOVA was conducted with stimulus set as a between-subject factor and probe type as a within-subject factor, there was no significant effect of stimulus set, and no significant interaction of stimulus set x probe type.

were excluded prior to the analysis, which caused 2.7% of the response time data from correctly responded trials to be dropped. No response was found to be quicker than 200ms. The means and standard errors of the trimmed response time data for each probe type were reported in Table 6.2. Paired t-test was conducted between participants' response time for visual common ground objects and privileged ground objects, and similarly, no significant difference was observed, $t$ (22) = .932, $p$ = .362, *Cohen's d* = .078.

One point worth noting is that the memory test in Experiment 9 consisted of 160 probe trials, while the memory test in Experiment 8 only consisted of 96 probe trials. A follow-up analysis was conducted to examine whether participants performed differently in the first half and the second half of the memory test due to fatigue effect in Experiment 9. Two-way repeated measures ANOVAs were conducted on participants' accuracy data and response time respectively, with probe type (visual common ground objects versus privileged ground objects) and probe order (the first 80 probes versus the last 80 probes) as two within-subject factors. For accuracy data, a main effect of probe order was observed, $F(1,22) = 26.499$, $p < .001$. Participants performed significantly better on the first 80 probe trials compared to the last 80 probe trials. No main effect of probe type was found, $F(1, 22) = 1.495$, $p = .234$, nor was there any interaction of probe order by probe type, $F(1, 22) = 1.080$, $p = .310$. As we can see from Table 6.3, for the first 80 probe trials, participants' average recognition rate of the visual common ground objects was higher than that of the privileged ground objects, whereas for the last 80 probe trials, the average recognition rates of the visual common ground objects and of the privileged ground objects seemed similar. For response time data, ANOVA revealed no main effect of order, $F(1, 22) = 2.110$, $p = .160$, no main effect of probe type, $F(1, 22) = .278$, $p = .603$, nor interaction between order and probe type, $F(1, 22) = 1.085$, $p = .309$.

Table 6.3. Means and standard errors of participants' accuracy and response time for each probe type, for the first 80 probe trials and last 80 probe trails in Experiment 9 respectively.

| | Accuracy | | | | | | Response Time (ms) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Visual common ground objects | | Privileged ground objects | | Foils | | Visual common ground objects | | Privileged ground objects | | Foils | |
| | Mean | SE | Mean | SE | Mean | SE | Mean | SE | Mean | SE | Mean | SE |
| First 80 trials | .637 | .030 | .584 | .029 | .957 | .011 | 980 | 44 | 995 | 44 | 1112 | 70 |
| Last 80 trials | .521 | .030 | .512 | .032 | .942 | .023 | 962 | 45 | 929 | 40 | 997 | 48 |

### 6.3.3. Discussion

No evidence was clearly observed to support our expectation that participants would show better memory for visual common ground objects compared to privileged ground objects. However, two non-significant trends are worth noting: first, the accuracy for visual common ground objects tended to be higher than the average accuracy for privileged ground objects, although not statistically significant; second, participants performed significantly worse on the last 80 probe trials than on the first 80 probe trials, and their performance on the first 80 probe trials revealed a tendency which was in line with our expectation. Based on these findings, we speculated that the duration of the memory test in Experiment 9 might have been too long so that participants' performance on the last half of the memory test was contaminated by fatigue. Given the positive results of Experiment 8, these more modest effects were further investigated in a follow-up experiment. In Experiment 10, the memory test contained a reduced number of memory probes, but the sample size was enlarged to make up the shortage of data points from each participant.

## 6.4. Experiment 10

### 6.4.1. Methods

#### *6.4.1.1. Participants*

Forty-two students (38 female, mean age 19.21, age range 18 to 22) from the University of Birmingham participated in this study in return for study credits. Twenty-one participants took the original version of the director task and correspondingly the stimulus set A of the memory task, the other twenty-one participants took the swapped version of the director task and the stimulus set B of the memory task. All participants had normal colour vision and normal or corrected-to-normal visual and auditory acuity. One additional participant participated in the study but did not complete it due to technical breakdown.

#### *6.4.1.2. Design and procedure*

The design and procedure were identical to that of Experiment 9 with the following exceptions. The length of the memory test was reduced. The memory test consisted 64 true probes and 32 foils. The number of true probes (including both visual common ground objects and privileged ground objects type) was reduced from 56 per type in Experiment 9 to 32 per type in Experiment 10, with 2 probes of each type from each shelf image. The number of foils also reduced from 48 to 32.

### 6.4.2. Results

**Director task**

Among the 42 participants, 1 participant committed 2 egocentric errors out of 8 critical instructions, 7 participants committed only 1 egocentric error each, and the other participants did not commit any egocentric errors. The overall accuracy rate is .973.

**Memory test**

Ten participants whose overall accuracy in the memory test failed to be significantly above chance level (i.e., failed to correctly response to 57 or more probes out of 96 probes) were excluded from the following analysis on both accuracy and response time.

Participants' performances on the memory test with the stimulus set A (16 participants) and stimulus set B (16 participants) were analysed together.[11] The means and standard errors of the accuracy data for each probe type were reported in Table 6.1. Paired t-test revealed no difference between participants' recognition memory for linguistic common ground objects and their recognition memory for visual common ground objects, $t(31) = .867$, $p = .393$, *Cohen's d* = .147.

For the response time analysis, only the trials to which participants responded correctly were included. Response times longer than 2 standard deviations from mean response time were excluded prior to the analysis, which caused 4.1% of the response time data from correctly responded trials to be dropped. No response was found to be quicker than 200ms. The means and standard errors of the trimmed response time data for each probe type was reported in Table 6.2. Paired t-test was conducted between participants' response time for visual common ground objects and privileged ground objects, and no significant difference was observed, $t(31) =1.050$, $p = .302$, *Cohen's d* = .132.

### 6.4.3. Discussion

The length of the memory test was reduced in Experiment 10 with the purpose of eliminating the potential fatigue effect. However, the findings from both accuracy and response time showed that participants were equally good at recognising the visual common ground objects and privileged ground objects during the memory test. No evidence was found

---

[11] When a mixed 2 x 2 ANOVA was conducted with stimulus set as a between-subject factor and probe type as a within-subject factor, there was no significant effect of stimulus set, and no significant interaction of stimulus set x probe type.

to confirm the hypothesis that participants would automatically form a better memory for objects which were visible to the director but not verbally referenced by the director, compared to objects which were occluded from the director's perspective. The implications of these findings and the potential limitations of the present study will be discussed with details in the next section.

## 6.5. General discussion

### 6.5.1. Summaries of Experiment 8-10

To our knowledge, the present study was the first to examine whether the status of an object in a conversational partner's perspective (i.e., common ground versus privileged ground) would implicitly influence participants' memory for these objects. Three experiments were conducted systematically using the same paradigm. Participants first participated in an adapted version of the director task, during which they followed the director's instructions to move some objects on a shelf. Participants then took an unexpected memory test with varied numbers of single recognition probes, during which they were asked to correctly recognise the objects which had been presented on the shelves in the previous director task and to correctly reject the foils. We distinguished three different types of objects presented on the shelf: linguistic common ground objects, visual common ground objects, and privileged ground objects. The present study was particularly interested in whether participants would automatically form stronger memory representations and, as a result, show better recognition memory for objects which could be seen but were not mentioned by the director (i.e., visual common ground objects), compared to objects which could not be seen by the director (i.e., privileged ground objects).

Experiment 8 showed that participants were more accurate at recognising the linguistic common ground object than the visual common ground objects, and also more accurate at recognising the visual common ground objects than the privileged ground objects, although the difference between the second pair of objects was not particularly strong. Experiment 9 was conducted with the aim of replicating the effect (and hopefully revealing a stronger effect) between visual common ground objects and privileged ground objects. However, no statistically significant difference was found between participants' memory for visual common ground objects and that for privileged ground objects. Participants performed significantly worse in the second half of the memory test compared to the first half of it. In order to ensure that the predicted effect was not hindered by fatigue, Experiment 10 was conducted with the length of the memory test being reduced. Findings from accuracy and response time showed that participants seemed to be equally good at recognising the visual common ground objects and the privileged ground objects. Overall, the three experiments in the current study failed to demonstrate stronger memory representations for visual common ground objects than for privileged ground objects.

### 6.5.2. Implications and limitations of the current findings

As mentioned in the introduction section, our first expectation was that participants would show more accurate memory for linguistic common ground objects compared to visual common ground objects. This expectation was confirmed by the finding from Experiment 8, supporting the idea that the conversational participants rely on the common ground representations whose nature is more likely to be 'gradient' and 'rich' rather than 'binary' and 'one-bit' (Brown-Schmidt, 2012). The current finding provides some preliminary evidence that people's memory of common ground information is likely to be affected by how common ground is established. Linguistic interaction, proposed by Clark and Marshall (1981) as one of the three bases on which common ground is established, might have increased the availability of the representations of common ground information in participants' memory on top of the basis of physical co-presence. However, it is also worth noting that the term ‘linguistic common ground objects’ was chosen for the current study to contrast with the term ‘visual common ground objects’, and to correspond to the ‘linguistic co-presence’ and ‘visual/physical co-presence’ concepts proposed by Clark and Marshall (1981). Although the common ground objects verbally mentioned by the directors in the current experimental paradigm were termed as ‘linguistic common ground objects’, these objects were not just talked about, but participants also looked at them for longer than at the other objects, and importantly, participants also physically interacted with them. Therefore, apart from the extra linguistic interaction, these objects also received more visual attention and physical interaction, which might have also contributed to participants’ higher recognition rates for these ‘linguistic common ground objects’ compared to the ‘visual common ground objects’. It would be interesting for future studies to tease these potential causes apart and to further investigate the richness of common ground representations established via different bases or via the combination of several basis, e.g., linguistic interaction, physical interaction, visual attention and physical co-presence etc. For example, a condition where participants are merely ask to follow the director’s instructions to fixate on the target object without making any physical interaction could be added to the current paradigm to distinguish the effect of physical interaction and the effect of linguistic interaction on memory.

Secondly, we expected participant to show better memory for objects in the common ground established by mere physical co-presence than objects in the privileged ground. We asked the question of whether participants would spontaneously pay more attention to information in common ground compared to information in privileged ground when we are engaged in communication. Our rationale was that if more attention was spontaneously assigned to common ground information compared to privileged ground information, then we would expect participants to show a better memory for the visual common ground objects compared to privileged ground objects. However, no clear difference was found between the visual common ground objects and privileged ground objects, which was the main focus of the present study. Overall, the present study did not lend support to the idea that our attention and memory, which are believed to be relatively low-level domain-general cognitive processes, may grant common ground information a prioritised status over privileged ground information.

One plausible explanation is that in every communicative situation, people do not necessarily and consistently pay more attention to and thus form stronger representations for information in common ground, compared to information in privileged ground. The way people assign attention to communicatively relevant information may be constrained by many context-specific factors, for example, the ratio between common ground information and privileged ground information, or people's communicative motivation. Supportively, the eye-tracking study reported in Chapter 2 showed that participants dwelled longer at each common ground object compared to each privileged ground object when there were fewer common ground objects on the shelf. In contrast, they dwelled longer at each privileged ground object than each common ground object when there were fewer privileged ground objects on the shelf. In the present experiments, there were always 4 objects in the privileged ground and 8 objects in the common ground. It is possible that the unequal ratio between common ground information and privileged ground information may have made the privileged ground objects more salient than common ground objects, and may have drawn participants' attention more towards privileged ground objects. One way to examine this possibility is to make the ratio between the number of common ground objects (including both linguistic common ground objects and visual common ground objects) and that of privileged ground objects equal. If participants start to show better memories for visual common ground objects than for privileged ground objects after such change, then it suggests that the lack of evidence for our hypothesis from the experiments in this chapter may be due to the confounding variable of the ratio between common and privileged ground information.

It is also plausible that the memory test used in the present study was unable to capture the effect we expected to find. In the current experiments, single recognition probes were used to assess participants' memory performances. In our everyday life, however, people are not usually explicitly asked to recognise whether a piece of information is shared between themselves and their conversational partners or not in a non-communicative setting. Free recall tests might have some advantages over recognition tests, but it also has some disadvantages. For example, as suggested by Mandler, Pearlstone, & Koopmans (1969), recognition typically examines the availability of information, while recall examines how accessible it is; and free recall is generally more difficult compared to recognition. It is unclear how participants' performance might be affected if a free recall task replaces the current recognition task. Another doubt is that, similar to recognition, people are usually not required to explicitly freely recall what information is shared between themselves and their conversational partners in the absence of any communicative cue. Perhaps a better suggestion would be to incorporate some communicative elements into the current memory test, so that the context of the memory test resembles our real-life communicative context under which people usually retrieve the common ground representations and privileged ground representations stored in memory. For example, the memory probes could occur in the presence of the director, as people usually retrieve relevant representations in the presence of a conversational partner. A further suggestion is to have memory probes presented following the director's voice cues alongside the presence of the director, as previous studies have demonstrated that listeners use speaker's voice as a cue to infer speaker's identity in order to efficiently retrieve relevant perspective representations during language comprehension (e.g., Ryskin, Wang, & Brown-Schmidt, 2016).

Another interesting finding was that participants' recognition rates for the visual common ground objects and privileged ground objects were close to chance level, while their accuracy rates for foils were relatively high. The high accuracy rates for foils indicate that participants' close-to chance level recognition rates for the positive probes (visual common ground and privileged ground objects) was not due to random selection. It is plausible that participants might not have paid close attention to all the objects presented in the grid during the director task, especially for the visual common ground and privileged ground objects that they had not linguistically and physically interacted with. Therefore, participants might have been biased to mark the objects that did not appear to be adequately familiar to them as 'false' in the memory test, resulting in their relatively high accuracy rates of correctly rejecting foils and low accuracy rates of correctly recognizing visual common ground and privileged ground objects. Future research could consider employing two-alternative forced-choice memory tests rather than single recognition probes to eliminate this type of response bias (e.g., Macmillan & Creelman, 1991). Another possible reason for this finding was that participants might have actively inhibited the representations of the visual common ground and privileged ground objects, which led to their low accuracy rates of these objects. Previous research has suggested that many instances of forgetting may result from active inhibitory processes which leads to the selective retrieval of relevant information (Anderson, 2001; Colzato et al., 2008), e.g., the retrieval-induced forgetting phenomenon (Anderson, Bjork, & Bjork, 1994). To test this explanation, future research could add a dual task condition where participants' active inhibitory resources are taxed by a secondary task alongside the main memory test. If the low recognition rates of visual common ground and privileged ground objects observed in the current study were due to active inhibition, then we would expect improved recognition rates for visual common ground and privileged ground objects.

### 6.5.3. Conclusion

In conclusion, the current study asked a novel question of whether the feature of objects being in the common ground versus privileged ground affected participants' memory for these objects, but the three experiments did not provide consistent evidence to answer this question. Nonetheless, the current study found that participants were better at recognising the objects in the common ground established by both linguistic and physical co-presence, compared with the objects in the common ground established by mere physical co-presence. This finding confirms that the nature of common ground representations in our everyday communication is likely to be graded and rich rather than binary. Future research is needed to examine what factors might constrain the richness of the common ground representations and privileged ground representations people form during communication. Additionally, it might be possible to explore a more suitable way to test participants' memory in order to address the research question we are asking in this chapter.

# CHAPTER VII

# General discussion

## 7.1. Overview

The aim of the present thesis was to examine the interplay between memory and common ground, more specifically, the role of long-term memory and working memory in forming, retaining and using common ground representations during communication. The thesis addressed this research question from two different angles. Firstly, Chapters 2 and 3 examined how participants' ability to use common ground information in language comprehension was constrained by the memory demands of the tasks and memory capacities of participants. Secondly, Chapters 5 and 6 explored whether the effects of common ground could be achieved via the low-level memory-based mechanism proposed by Horton and Gerrig (2005a, 2005b), which was introduced in Chapter 4.

In the following sections, I will first summarise the main empirical findings from the thesis. I will then critically discuss the overall findings and their theoretical implications by grouping evidence into two themes, i.e., the sensitivity to common ground during communication, and the role of memory in the use of common ground. Thirdly, I will discuss some future research directions opened up by the current work. Finally, I will conclude with the contribution of the current work to this research area.

## 7.2. Summary of empirical findings

Experiment 1 in Chapter 2 adapted a widely-used referential communication task (i.e., director task) to capture the memory demand of having to hold in mind someone's perspective for later language comprehension. Experiment 1 also manipulated memory load by systematically altering the relative size of common ground versus privileged ground, as well as measuring individuals' working memory capacities. With the adapted director task, the present study succeeded in finding distinct memory-related effects during perspective encoding and during the integration of perspective information with the director's instructions. First, participants' eye-movement records during the perspective encoding phase were sensitive to the relative size of common versus privileged ground, and participants with medium and high working memory capacities showed a greater sensitivity to the relative size of common versus privileged ground. Second, during the integration of perspective and instructions, participants committed more egocentric errors as the size of the common ground increased. More egocentric errors were also observed among participants with low working memory capacities, which is consistent with the suggestion that integration of perspective and instructions draws heavily upon domain-general working memory resources (Lin et al., 2010). However, these two main effects of common ground size and working memory capacity did not interact with each other. Therefore, I argue that it is plausible that common ground size and working memory capacity factors play dissociable roles in perspective encoding and perspective integration respectively. Common ground size might have influenced participants' egocentrism via the encoding and retaining process of long-term memory, whereas working memory might be more heavily engaged in integrating perspective information with verbal messages.

Built upon the foundations laid by Chapter 2, Chapter 3 took a developmental perspective to examine the relationship between memory and perspective use in school-aged children. As reviewed in Chapter 3, acquiring conceptual 'theory of mind' understanding is not adequate for successfully using a conversational partner's perspective information in language comprehension (e.g., Keysar et al., 2000, 2003), and children's ability to use a speaker's perspective undergoes a prolonged developmental trajectory until adulthood (e.g., Dumontheil et al., 2010). It is therefore important to examine what memory-related factors affect children's ability of using perspective in online communication, and whether the same factors contribute to their age-related development in perspective use. In experiments 2 and 3, the memory demand of holding in mind a speaker's perspective was enforced in the hidden perspective task version and was minimal in the visible perspective task version. The size of common ground was manipulated in two different ways in Experiment 2 and Experiment 3 respectively. The two experiments examined the effects of these two memory demands on 8- and 10-year-old children's perspective-taking performance during language comprehension. While 10-year-olds were significantly less egocentric than 8-year-olds in Experiment 2, this age-related improvement was marginal in Experiment 3. Both 8- and 10-year-olds committed more egocentric errors when they were under the enforced memory demand of holding in mind a speaker's perspective than when this demand was minimal, and were more egocentric when they had to manage a larger amount of information in common ground. However, as discussed in Chapter 3, no clear evidence was found that these two memory factors were the factors that drove age-related improvement in perspective use, as no significant interaction was observed between age and these two memory factors in Experiment 3 (after controlling for a confound variable in Experiment 2).

A series of experiments presented in Chapter 5 examined whether the change of conversational partners during communication would introduce an event boundary, thus making recently encoded objects less accessible in memory, compared to when there was no such change. Two different experimental paradigms were employed in Chapter 5 to seek evidence for this proposed low-level memory-based mechanism for partner-specific object representations. Experiment 4 provided evidence showing that participants' memory for the recently encoded objects did become less available after experiencing a change of conversational change; however, Experiments 5, 6 and 7 failed to consistently replicate this partner-specific effect. These findings were only tentative, but they casted doubt on the idea of a low-level memory-based mechanism underlying common ground which was thought to be automatic (Horton & Gerrig, 2005a, 2005b). There is no doubt that people do form and retrieve common ground representations in and from memory to guide their communication.

However, findings from Chapter 5 suggest that this memory-based mechanism for common ground might not be completely automatic and stimulus-driven, but instead might be constrained by other social or cognitive factors, for example, the communicative relevance of a partner in a conversational context.

Chapter 6 used a newly-adapted version of the director task to explore a novel question—whether the feature of objects being in the common ground versus privileged ground in a preceding discourse influenced participants' recognition memory for these objects in a later unexpected memory test. If participants spontaneously paid more attention and thus form stronger memory representations for information in the common ground compared to information in the privileged ground, then they would be expected to show a better recognition memory for objects in the common ground. Unfortunately, Experiments 8 to 10 failed to provide consistent evidence to support this hypothesis. One possible reason for this failure is that the ratio between the numbers of objects in the common ground versus privileged ground may have directed participants' attention more towards objects in the privileged ground. However, this finding also opens up the possibility that people do not necessarily form and retain stronger representations for common ground compared to privileged ground.

## 7.3. Discussion of general findings

### 7.3.1. Sensitivity to common ground during communication

As discussed in Chapter 1, the existing literature demonstrates that people do show sensitivity to common ground during communication. However, what remains controversial is the questions about when, how and to what extent people show sensitivity to common ground in language processing. Four different accounts regarding this debate have been reviewed in Chapter 1, namely, the reference diary account (Clark & Marshall, 1978), perspective adjustment model (Epley, Keysar, et al., 2004; Keysar et al., 2000, 1998), constraint-based model (e.g., (Hanna et al., 2003), and anticipation-integration model (Barr, 2008). In this section, I will review the present findings regarding the sensitivity to common ground in this thesis, discuss how the empirical findings fit with these different theoretical accounts, and then further propose that successful communication needs to take into account both common and privileged ground information.

#### *7.3.1.1. How the current findings fit with different accounts of common ground*

Chapters 2 and 3 examined participants' sensitivity to common ground information from two aspects, i.e., their sensitivity to common ground during perspective encoding, and their use of common ground during perspective integration. On the one hand, chapter 2 showed that adult participants were sensitive to the size of common ground during perspective encoding. Moreover, such sensitivity to the size of common ground during perspective encoding was not modulated by participants' high or low working memory capacities. This finding indicated that participants were aware of the need to take common ground into account, and were efficient at distinguishing and registering what was in common ground and what was in privileged ground regardless of their working memory capacities. On the other hand, findings from Chapter 2 and Chapter 3 suggested that both adults and children had difficulty in using common ground information during perspective integration despite their awareness of the need to take common ground into account. This difficulty in the use of common ground was greater when there was a larger amount of common ground information.

First, the current findings have provided some evidence which cannot be fully explained by the reference diary account (Clark & Marshall, 1978). The reference diary account suggests that in every communicative situation people search in memory for events which allow them to know if certain information is in common ground. According to Clark and colleagues, common ground serves as a primary context which is central to language, and it fully constrains language comprehension by defining the relevant domain of interpretation (Clark, 1992; Clark & Wilkes-Gibbs, 1986). However, the current findings demonstrate that common ground did not serve as a full constraint over participants' language comprehension, as participants still erroneously selected competitors in the privileged ground in the current experiments. Admittedly, the egocentric errors in the hidden perspective version could be due to memory failure rather than failure in assessing common ground, as participants memory for what objects were in the common ground might be erroneous. However, this memory failure account would have difficulty in explaining why participants still frequently made egocentric errors even when they were not required to remember relevant perspective information in mind (e.g., Apperly et al., 2010; Keysar et al., 2000; and also findings from the visible perspective version in Experiment 2&3). More importantly, the cognitive mechanism of the reference diary account was unspecified by Clark & Marshall (1978). Therefore, although the evidence observed in the current study does not necessarily rule out the reference diary account, it does not support the reference diary account either.

The remaining three accounts all acknowledge that common ground does not fully constrain language comprehension, but the accounts differ on the view of whether common ground can immediately constrain language comprehension. The constraint-based account (Hanna et al., 2003) views language comprehension as a single, integrated, probabilistic and interactive process in which common ground functions as a constraint alongside other potential constraints. This constraint-based account suggests that common ground can immediately constrain language processing. The perspective adjustment account (e.g., Keysar et al., 2000) and anticipation-integration account (Barr, 2008) both view language comprehension as a two-stage process and both emphasise the role of common ground in the later correction phase. One difference between the perspective adjustment account and anticipation-integration account is that the former only recognises people's use of common ground in a late “adjustment” stage, whereas the latter not only recognises their use of common ground in a late “integration” stage, but also recognises their sensitivity to common ground in an early ‘anticipation’ stage where no perspective use is required. Another difference is that the former regards the egocentric effect as an adaptive strategy or default which is generally beneficial for social information processing, whereas the latter regards the egocentric effect as an autonomous activation process which is beyond control and even awareness (see Barr, 2008 for a more detailed comparison between these two accounts).

The perspective adjustment account would have difficulty accounting for the finding that participants showed considerable sensitivity to the size of common ground during perspective encoding in Chapter 2. According to perspective adjustment account, people only show sensitivity to common ground during the later adjustment phase but not the early phase, which is not the case in Chapter 2. In contrast, the anticipation-integration account could potentially explain participants’ sensitivity to common ground during perspective encoding as participants anticipated referents to be from common ground, although participants in Chapter 2 did not show strong preference for common ground prior to perspective integration as observed in Barr (2008). The fact that adult participants still made egocentric errors despite their awareness of the need to accommodate the speaker’s limited perspective is also consistent with the idea from anticipation-integration account that people suffer from the autonomously activated interference of privileged information during the early moments of perspective integration. The finding that adults with higher working memory capacities made less egocentric errors also accords with the idea that the later correction process in which common ground information plays a role draws upon people’s executive functions capacities. People with higher working memory capacities might be more capable of correcting the initially activated egocentric bias.

Lastly, although the current findings do not directly provide evidence in favour of the constraint-based account, the findings do not contradict this account. The constraint-based account could explain the sensitivity to common ground during perspective encoding as a result of common ground information acting as a salient constraint before the onset of instructions. It could also explain the varied degree of egocentrism during perspective integration as a result of competition between linguistic constraints (e.g., 'the large hat') and pragmatic / perspective constraint (e.g., the fact that the largest hat is blocked from the director's perspective). It is plausible that when linguistic constraints sometimes overrode pragmatic constraints, egocentric errors would occur. Overall, the experiments in the thesis have contributed some evidence to the debates on when and how people show sensitivity to common ground during language processing.

#### *7.3.1.2. Successful communication requires consideration of both common ground and privileged ground*

As summarised in Chapter 4, consideration of common ground is often related to the concepts of 'audience design', 'perspective taking' and 'partner-specific adaptation' in the existing literature. However, attention to privileged ground or the use of privileged knowledge is often labelled as 'egocentrism', 'failure of perspective taking' or 'miscommunication'. The current findings demonstrate that paying attention or showing sensitivity to privileged ground itself is not the same thing as 'egocentricity'. For example, Chapter 2 found that over half of the participants explicitly adapted their eye-movements based on the relative size between common ground versus privileged ground, which suggests that people may choose to attend to privileged ground information in order to remind themselves not to move objects in privileged ground. Moreover, participants who explicitly reported their use of such a memory strategy and those who did not were equally good at using perspective information to identify the target referent. This suggested that paying more attention to privileged ground information did not necessarily lead to egocentrism. Chapter 6 further confirmed that both common ground information and privileged ground information received a considerable amount of attention during communication from participants who successfully used the director' perspective information in most critical trials (most participants made no egocentric errors and only a few participants each made one egocentric error out of eight critical trials). Experiments in Chapter 6 provided no evidence in support of the hypothesis that participants' memory traces for objects in visual common ground were any stronger compared to their memory traces for objects in the privileged ground. Altogether, the evidence from the present thesis clearly suggests that people do not only form common ground representations but also form privileged ground representations during communication, and paying attention to privileged ground does not necessarily lead to egocentrism.

Successful communication requires people to distinguish shared knowledge from privileged knowledge and to use the appropriate set of knowledge from common ground or/and privileged ground, rather than merely drawing upon common ground information. However, most research has focused on the processes underlying the formation and use of common ground representations in communication, and there has not been much research investigating when and how people may form and use privileged ground representations in communication. A study by Lane et al. (2006) showed that participants did not only have difficulty avoiding using privileged knowledge, but that reminding oneself not to mention the privileged information could even make it harder to guard against giving the privileged information away. This study focused on people's use of privileged knowledge, however it was still designed under the notion that the use of privileged knowledge should be avoided. Another research by Brown-Schmidt et al. (2008), which was reviewed in Chapter 4, showed that listeners rapidly directed their attention toward items in their privileged ground when they heard instructions in the form of questions rather than of imperatives. This finding implies that conversational participants may interpret a speaker's use of questions as an indication of the speaker's lack of certain information, which guide their attention towards privileged rather than common ground information. This type of sensitivity to privileged ground reflects people's understanding of others' mental states rather than reflecting one's egocentricity.

It would be a fruitful area for future researchers to design experiments or adapt the existing paradigms in order to address questions regarding the formation and use of privileged ground representations as well as those of common ground representations. For example, when do people prioritise their attention to information in privileged ground over information in common ground, and how do people use both privileged ground representations and common ground representations to guide communication.

## 7.3.2. The role of memory in common ground

### *7.3.2.1. The role of memory in perspective taking during communication*

In Chapter 1, we reviewed the theories of the memory processes involved in language comprehension, e.g., working memory, long-term memory and potentially long-term working memory. However, as discussed in the introduction section in Chapters 2 and 3, a lot of research using communication tasks has examined the relationship between working memory and perspective taking (Cane et al., 2016; Lin et al., 2010; Nilsen & Graham, 2009; Ryskin et al., 2015; Wardlow, 2013), but very little is known about the role of long-term memory in perspective-taking processes. The experiments in Chapters 2 and 3 pioneered a way to incorporate the memory demand of holding in mind a speaker's perspective into a communication task, and to examine how this memory retention load contribute to the use of online visual perspective taking. Apart from enforcing the memory load of retaining a speaker's perspective for seconds, the current experiments also systematically manipulated the number of common versus privileged items. Additionally, Chapter 2 measured individuals' working memory capacities using the OSPAN task. In this section, I will discuss how these memory-related manipulations might inform us of the roles of working memory and long-term memory in perspective taking, and also in the development of perspective taking in children.

In terms of the role of working memory, Chapter 2 provided evidence that working memory was heavily involved in perspective integration phase. The lower one's working memory capacity, the fewer memory resources one could utilize to integrate perspective information into communication, and the more egocentric errors one would commit. It was speculated that the demand of managing an increasing common ground size might have taxed adult participants' long-term memory rather than working memory.

In order to understand the role of long-term memory in perspective taking, two questions are important to address: 1) if the demand of having to remember a speaker's perspective prior to perspective integration had any effect on participants' overall performance; 2) if the demand of having to remember a speaker's perspective modulated the effect of the size of common ground on egocentrism. The first question addresses whether the longer-term memory encoding and retention loads affected perspective taking. The second question is interested in whether the effect of common ground size interacted with the effect of enforced encoding and retention load. An interaction would potentially indicate that the demand of managing an increasing size of common ground information taxed participants' long-term memory resources.

It is worth noting that Experiment 1 in Chapter 2 only employed the hidden perspective task version which enforced the memory load of remembering the director's perspective, but had no control condition which resembled the visible perspective version in Chapter 3. Nonetheless, a previous study by Wang et al. (submitted), which manipulated the relative size of common ground versus privileged ground in a visible perspective version of the director task, may be able to shed some light on these two questions. One experiment in Wang et al. (submitted) systematically manipulated the relative magnitude of common ground versus privileged ground in a classic director task by varying the number of blocked slots on the shelf (3, 5, 7 or 9 slots) while keeping the total number of presented objects constant as 12. Overall, participants committed fewer egocentric errors in Wang et al.'s study (ranging from 2.1% to 5%) than in Chapter 2 (ranging from14.6% to 31.1% across different magnitude groups). More importantly, results from Wang et al.' s study showed that the magnitude factor had no systematic effect on adults' perspective taking performance in the director task. A summary of the findings from Wang et al. (submitted) and Experiments 1-3 of the thesis can be found in Table 7.1.

The comparison between findings from Chapter 2 and findings from Wang et al. (submitted), while preliminary, may help us answer the two aforementioned questions for adult participants. [12] As to the first question, this comparison indicates that the demand of having to remember the director's perspective had a detrimental effect on adult participants' overall performance. The enforced memory demand made it generally more difficult to successfully encode, maintain or use the person's perspective information in communication. The variation of this memory demand across different communicative situations could partially account for the variation in people's performances of using their communicational partner's perspective during language processing. Second, it seems that the size of common ground only affected adult participants' egocentrism when it was necessary to hold in mind the speakers' appropriate perspective information for successful language comprehension.

[12] Please note that Wang et al. (submitted) manipulated the number of blocked slots on the grid, but there were some blocked slots which did not contain any object. However, Experiments 1-3 in this thesis all manipulated the number of objects in the privileged ground. Therefore, we could not conduct direct comparison between the findings from these two studies.

When the demand of having to remember the director' perspective reduced to minimal, adults might be generally able to efficiently and effectively track and use the director's perspective online during communication. Previous research showed that participants were efficient at online calculation of simple level-1 perspective which concerned whether another person could see a certain item or not (Qureshi et al., 2010; Samson et al., 2010). Such efficient calculation is likely to be a relevant process in solving the director task with minimal retention load. This finding further supports the conclusion in Chapter 2 that the size of common ground may have affected the rates of egocentric errors via a long-term memory process. Working memory capacities, on the other hand, may have directly affected the likelihood of successfully integrating the appropriate perspective information with the speaker's message.

Do working memory and long-term memory contribute to children's perspective-taking behaviours similarly as to adults'? Chapter 3 provides direct evidence to answer the aforementioned two questions. Firstly, both 8- and 10-year-olds made significantly more egocentric errors in the hidden perspective version than in the visible perspective version. This finding further confirms the idea that the memory demand of remembering what the speaker knows, which is commonly embedded in everyday communication, constrains how well adults and children can use perspective information to guide communication. Secondly, in Chapter 3, a main effect of the size of common ground on egocentrism was observed in both 8- and 10-year-olds: the larger the common ground size was, the more egocentric errors the children made. However, no clear evidence was found that this main effect of the size of common ground on egocentrism interacted with the effect of task version in Experiment 3, after controlling for the confounding variable in Experiment 2. This result indicated that the size of common ground affected 8- and 10-year-olds' degree of egocentrism even when it was unnecessary for them to remember the perspective information.

The second finding seems inconsistent with the conclusion from Chapter 2 that the size of common ground may have affected adult participants' rate of egocentric errors via a long-term memory process. One possibility is that whereas the size of common ground may have affected adults' perspective use mainly via a longer term memory route, it may have also affected children's perspective use via a working memory route. When under the minimal load of remembering the speaker's perspective, adults may be generally able to track and also use the director's perspective online during communication efficiently and effectively. Children may be as efficient as adults at calculating the speakers' perspective online in the director task, as Surtees and Apperly (2012) found that children and adult participants were both efficient at calculating simple level-1 perspective which concerned whether another person could see a certain item or not. However, it is plausible that the demand of integrating an increasing size of common ground with verbal messages may considerably tax children's cognitive resources (especially working memory), as their executive functions are still developing and their social experience is rather limited. One way to test this possibility is to add the measurement for children's working memory capacities (e.g., OSPAN task) into the current developmental experiments, and examine whether ANOVA with working memory capacities and the size of common ground as two factors would reveal a significant interaction. If working memory capacity modulates the effect of common ground size on children's rates of egocentrism, then this possibility can be confirmed.

Table 7.1. A summary of the findings from Wang et al (submitted), Chapter 2 and Chapter 3.

| | Factors examined | Key manipulation | Key findings: Perspective encoding phase | Key findings: Perspective Integration phase |
|---|---|---|---|---|
| **Wang et al. (submitted)** | 1. Visible Perspective version<br>2. The relative size of common ground versus privileged ground | 3, 5, 7, or 9 blocked slots while the total number of objects was held at 12 | Longer dwell time on objects in the common ground compared to those in the privileged ground (expect for the condition with 5 blocked slots) | • Egocentrism was observed in proportional egocentric errors (as participants made more egocentric errors in the experimental condition than in the control condition). Egocentrism was not observed in response time or latency to final target fixation (as participants did not respond more slowly in the experimental condition than in the control condition).<br>• The proportional egocentric errors did not serve as a function of the object array size. |
| **Chapter 2 Experiment 1** | 1. Hidden perspective version<br>2.The relative size of common ground versus privileged ground<br>3. Working memory capacity (OSPAN task) | 3, 5, 7, or 9 objects in privileged ground while the total number of objects was held at 12 | The smaller the size of common ground was, the longer participants dwelled on the objects in the common ground,. | • Egocentrism was observed in proportional egocentric errors. Egocentrism was not observed in response time or latency to final target fixation.<br>• A systematic effect of the size of common ground: the larger the size of common ground, the more egocentric errors one made.<br>• A main effect of working memory capacity: the higher one's working memory capacity was, the fewer egocentric errors one made.<br>• No interaction between the effects of common ground size and working memory capacity on egocentrism |
| **Chapter 3 Experiment 2** | 1. Age (8 &10)<br>2. Object array size<br>3. Task version (visible perspective version & hidden perspective version) | 5, 7, or 9 objects in total while the number of privileged objects was held at 2 | | • Egocentrism was observed in proportional egocentric errors.<br>• A main effect of age: fewer egocentric errors in 10-year-olds than in 8-year-olds.<br>• A main effect of task version: fewer egocentric errors in visible perspective version compared to hidden perspective version<br>• A main effect of the size of common ground: The larger the common ground size, the more egocentric errors one made.<br>• A significant 3-way interaction of age x task x array size. |
| **Chapter 3 Experiment 3** | 1. Age (8 &10)<br>2. The relative size of common ground versus privileged ground<br>3.Task version (visible perspective version & hidden perspective version) | 3, 5, 7, or 9 objects in privileged ground while the total number of objects was held at 12 | | • Egocentrism was observed in proportional egocentric errors.<br>• A marginally significant effect of age: fewer egocentric errors in 10-year-olds than 8-year-olds.<br>• A main effect of task version: : fewer egocentric errors in visible perspective version compared to hidden perspective version<br>• A main effect of the size of common ground: The larger the common ground size, the more egocentric errors one made.<br>• A marginally significant interaction of age x task version |

Overall, the present thesis proposes that working memory may be important for integrating common ground information with the speaker's current utterance over short periods of time and with a storage capacity limit of just a small number of items. In contrast, long-term memory may be the main resource for retaining common ground information over an extended discourse. This idea is consistent with the literature on the memory process involved in text comprehension, which was reviewed in Chapter 1 (Ericsson & Kintsch, 1995; Graesser et al., 1997).

#### *7.3.2.2. The low-level memory-based mechanism for common ground*

The traditional view of common ground by Clark and Marshall (1978, 1981) suggests that individuals mark their memory for an event with information about who is present during the event. They also explicitly search in their memory for signs of co-presence to assess whether a piece of information is shared with someone. Horton and Gerrig (2005a, 2005b) have developed their theory on the foundation of this traditional view that common ground information is represented in long-term episodic memory. They further propose that under some situations people may rely on an automatic associative memory mechanism to efficiently retrieve and use common ground information during communication, without necessarily employing an explicit process to determine mutual knowledge which calls upon higher-order mindreading abilities. This low-level memory-based approach did not rule out the route of explicit mentalising to common ground, but it offers an alternative way of how the effects of common ground might be achieved with the support of a domain-general memory process, corresponding the 'submentalising' concept proposed by Heyes (2014),

Chapters 5 and 6 examined whether a change of conversational partners (Chapter 5) and the ground status of an object in a conversational partner's perspective (Chapter 6) would affect participants' memory for objects in a relatively automatic way. The results from Chapters 5 and 6 failed to show effects which could support the account of the low-level memory mechanism of common ground. While acknowledging the potential limitations of the current experimental designs (which have been discussed in sections 5.8 and 6.5 of this thesis), this failure to find predicted effects warrants some cautious discussion of the possibility that the memory-based mechanisms underlying the effects of common ground may not be completely automatic. A review paper by Brown-Schmidt et al. (2015) suggests that the *formation* of common ground representations may be constrained by cognitive factors (e.g., attention, memory) and contextual factors (e.g., communicative relevance). A further speculation is that the *retrieval* of common ground representations in communicative situations may also be constrained by these cognitive and contextual factors, and also social factors (e.g., moods, motivations, goals). It is undeniable that our memory for objects could be influenced by the presence/absence of someone whom those objects are associated with, and by whether the object is visible or invisible from someone's perspectives, but what remains unclear is under what circumstances these influences might occur. The suggestion of an automatic memory-based mechanism of common ground needs further careful investigation.

## 7.4. Future directions

There are a number of ways in which the current work can be developed in further research. First, it would be informative to conduct an eye-tracking version of Experiment 3 with children. The application of eye-tracking technique on children will allow us to develop a better understanding of the time course of children's sensitivity to common ground during perspective encoding and perspective integration. It would be particularly interesting to examine whether children in the hidden perspective version would show a similar pattern of eye movements during the perspective encoding phase as adults did in Chapter 2. If yes, then it would indicate that children are also sensitive to common ground from as early as the perspective encoding phase, and that children are efficient enough at encoding common ground and privileged ground information. It would also be interesting to examine whether children would show different patterns of eye movements in the visible perspective version and the hidden perspective version. The application of eye-tracking method on children would help develop a full picture of how children infer and use common ground information in different sub-processes of perspective taking. It will also help draw similarities and contrasts between the time courses by which common ground is incorporated in communication in adults and in children.

Second, compared to applying eye-tracking methods on both adults and children, a potentially cheaper and less time-consuming way is to use mouse-tracking or motion-tracking instead. The mouse tracking method has been proven to be useful in studying how people track their own and others' beliefs in recent theory of mind literature (e.g., van der Wel, Sebanz, & Knoblich, 2014). During the testing phase of the current experiments in Chapters 2 and 3, it was interesting to observe that participants, especially children, often approached a distractor in the privileged ground at the beginning, but then stopped half way and managed to reach towards the target referent in the common ground at the last moment. The trajectory of this 'half-movement' has the potential to offer a continuous measure of participants' degree of egocentrism, and may reveal some interesting findings which have been missed in the current results.

Third, in Chapters 2 and 3 where the role of memory in perspective taking was examined, the memory demands of the director task was primarily manipulated, whereas the measure of individual differences in working memory capacity is only limited to the use of OSPAN task on adults. OSPAN task is generally believed as a reliable and valid measure of working memory capacity that correlates to a wide variety of other cognitive tasks (e.g., Engle, Cantor, & Carullo, 1992; Unsworth & Engle, 2007; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). However, future research should extend its focus from working memory solely to other memory processes as well, e.g., long-term memory. An improvement could be made to the current design by using a battery of memory tasks (including working memory tasks and also long term memory tasks) to test both adults and children's individual differences in different aspects of memory processes, and to examine how each aspect of memory processes is involved in different sub-processes of perspective taking. The present findings from Chapter 2 imply that long-term memory and working memory might play dissociable roles in perspective encoding and perspective integration respectively. By adding the individual difference measure on longer term memory (e.g., cued recall test), future research can examine whether one's long-term memory capacity affects their perspective-taking performance under the enforced memory load of remembering a speaker's perspective. It can also examine whether the effect of the size of common ground interacts with the effect of long-term memory capacities on adults' degree of egocentrism.

Fourth, another important avenue is to delineate the nature of the common ground representations established via different ways. As reviewed in Chapter 1, common ground can be established physically, linguistically, or culturally (Clark & Marshall, 1978, 1981). Some researchers propose that the nature of common ground representation can be simply one-bit and binary—either in common ground or in privileged ground (Brennan & Hanna, 2009; Galati & Brennan, 2010). In contrast, Brown-Schmidt (2012) argues that common ground representations are likely to be gradient and rich. The present findings from Experiment 8 indicate that the richness and strength of common ground representations in memory can be influenced by how common ground is established, thus providing some evidence in support of Brown-Schmidt (2012)'s suggestion. However, most existing studies on perspective taking and common ground have only manipulated the status of common ground as a simple binary feature, and such binary nature has usually been established by only one single manipulation which is most likely to be visual / physical co-presence (e.g., the director task). In contrast, in our everyday life, we talk about the items which we are looking at, we direct other people's attention to the items we are talking about, we judge each other's communality membership by the things we read or talk about, and we use our culture knowledge to help us understand the things we talk about. The three bases of common ground, i.e., physical co-presence, linguistic co-presence and culture co-presence, hugely interact with each other. Future research could be done to examine how these different routes interplay with each other to shape the representations of common ground. I speculate that the representations of common ground established by a combination of more than one type of co-presence would be stronger and richer than the representations of common ground established by a single type of co-presence.

Lastly, another avenue for future research is to use a more interactive communication task and to examine the role of memory in common ground during both language comprehension and language production. Communication is a two-way interaction during which conversational participants take turns to be speakers and listeners, and it involves both language comprehension and language production. During such an everyday interactive communication, speakers and listeners also give feedback to each other's messages or interpretations, and also make adaptations to language comprehension and production accordingly. However, most studies examine either language comprehension or production in a one-way manner where speakers are always speakers and listeners are always listeners. Taking such an isolated approach to studying communication would reduce participants' motivation to take their conversational partner's perspective into account. (e.g., Yoon, Koh, & Brown-Schmidt, 2012). Additionally, memory might play different roles in the use of common ground during language comprehension versus language production. For example, research showed that speakers tended to have better memory for the content of a conversation than listeners (e.g., Yoon, Benjamin, & Brown-Schmidt, 2016). It is, therefore, important to study the role of memory in perspective taking during both language comprehension and language production in order to develop a full picture of this research question. One straightforward step following this thesis is to operationalise a way to incorporate the demand of remembering a listener's perspective and the demand of managing increasing common ground size into a language production task and to examine the effect of these memory-related factors on language production. For example, a previous study found that 4-year-old children tended to produce more complex expressions to listens when the object array size was larger in a referential communication task, compared to when there were fewer distractors in visual scenes (Matthews et al., 2012). Future research could further investigate how common ground size might affect both adults and children' perspective-taking performance in language production when they are under the memory demand of remembering a listener's perspective. A more advanced step could be to study how conversational participants use the perspective information they have previously encoded in memory in an interactive communicative scenario, where a single participant plays both the role of a listener and a speaker at various points.

## 7.5. Conclusion

In conclusion, the current thesis examined the role of memory involved in encoding, retaining and using common ground information shared between conversational participants during communication. This thesis converges on many areas, including attention, memory, social cognition, developmental psychology and psycholinguistics. Ten experiments presented across four empirical chapters in this thesis address the two main research questions proposed at the outset. One question is how memory constrains both adults' and 8- and 10-year old children's abilities to use a speaker's perspective information during language comprehension. The current work in Chapters 2 and 3 highlights the role of both working memory and long-term memory in perspective use. The other questions is whether the effects of common ground can be established with the support of a low-level memory-based mechanism, without necessarily going through high-level mindreading inference. The current work in Chapters 5 and 6 suggest that the memory-based mechanisms underlying the effects of common ground may be constrained by cognitive factors (e.g., attention, memory) and contextual factors (e.g., communicative relevance). I hope the methodologies and findings in this thesis will open up new avenues for future investigation on the role of memory underlying the effects of common ground during communication from different research angles and on different subject groups, in order to develop a full picture of how memory constraints and supports the use of common ground in our everyday communication.

## References


Abu-Akel, A. M., Wood, S. J., Hansen, P. C., & Apperly, I. A. (2015). Perspective-taking abilities in the balance between autism tendencies and psychosis proneness. *Proc. R. Soc. B*, *282*(1808), 20150563. https://doi.org/10.1098/rspb.2015.0563

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Anderson, M. C. (2001). Active forgetting: Evidence for functional inhibition as a source of memory failure. *Journal of Aggression, Maltreatment & Trauma*, *4*(2), 185–210.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217. https://doi.org/10.1080/17470210903281582

Astington, J. W., & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology*, *9*(1), 7–31. https://doi.org/10.1111/j.2044-835X.1991.tb00859.x

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. D. (1986). *Working memory / Alan Baddeley.* Oxford: Clarendon.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/S0079742108604521

Ballesteros, S., Mayas, J., Prieto, A., Toril, P., Pita, C., Reales, J. M., & Waterworth, J. A. (2015). A randomized controlled trial of brain training with non-action video games in older adults: results of the 3-month follow-up. *Frontiers in Aging Neuroscience*, *7*, 45.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40. https://doi.org/10.1016/j.cognition.2008.07.005

Barr, D. J. (2014). Perspective Taking and Its Impostors in Language Use: Four Patterns. *The Oxford Handbook of Language and Social Psychology*, 98.

Barr, D. J., & Keysar, B. (2002). Anchoring Comprehension in Linguistic Precedents. *Journal of Memory and Language*, *46*(2), 391–418. https://doi.org/10.1006/jmla.2001.2815

Blain-Brière, B., Bouchard, C., & Bigras, N. (2014). The role of executive functions in the pragmatic skills of children age 4–5. *Frontiers in Psychology*, *5*.

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.

Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*, *1*(2), 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Bristow, D., Rees, G., & Frith, C. D. (2007). Social interaction modifies neural response to gaze shifts. *Social Cognitive and Affective Neuroscience*, *2*(1), 52–61.

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190. https://doi.org/10.1016/j.jml.2009.04.003

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*(5), 893–900. https://doi.org/10.3758/PBR.16.5.893

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*(1), 62–89. https://doi.org/10.1080/01690965.2010.543363

Brown-Schmidt, S., & Duff, M. C. (2016). Memory and Common Ground Processes in Language Use. *Topics in Cognitive Science*, *8*(4), 722–736.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–1134. https://doi.org/10.1016/j.cognition.2007.11.005

Brown-Schmidt, S., & Horton, W. S. (2014). The influence of partner-specific memory associations on picture naming: A failure to replicate Horton (2007). *PloS One*, *9*(10), e109035.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as Contexts in Conversation. In *Psychology of Learning and Motivation* (Vol. 62, pp. 59–99). Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0079742114000048

Cane, J. E., Ferguson, H. J., & Apperly, I. (2016). Using perspective to resolve reference: the impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Retrieved from https://kar.kent.ac.uk/57114/

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*(2), 73–92.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*(4), 299–319.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. Guilford Press.

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

Clark, H. H., & Marshall, C. (1978). Reference Diaries. In *Proceedings of the 1978 Workshop on Theoretical Issues in Natural Language Processing* (pp. 57–63). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.3115/980262.980272

Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In *Elements of Discourse Understanding*. Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., & Hommel, B. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 302.

Conty, L., Tijus, C., Hugueville, L., Coelho, E., & George, N. (2006). Searching for asymmetries in the detection of gaze contact versus averted gaze under different head views: a behavioural study. *Spatial Vision*, *19*(6), 529–545.

Davis, H. L., & Pratt, C. (1995). The development of children’s theory of mind: The working memory explanation. *Australian Journal of Psychology*, *47*(1), 25–31.

Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*(2), 331–338. https://doi.org/10.1111/j.1467-7687.2009.00888.x

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: a test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 972.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective Taking as Egocentric Anchoring and Adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*(6), 760–768. https://doi.org/10.1016/j.jesp.2004.02.002

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Ferguson, H. J., Apperly, I., Ahmad, J., Bindemann, M., & Cane, J. (2015). Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task. *Cognition*, *139*, 50–70.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99.

Frick, A., Möhring, W., & Newcombe, N. S. (2014). Picturing perspectives: development of perspective-taking abilities in 4-to 8-year-olds. *Frontiers in Psychology*, *5*, 386.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*(4), 483–527.

Galati, A., & Brennan, S. E. (2006). Given–new attenuation effects in spoken discourse: For the speaker, or for the addressee. In *Psychonomic Society, 47th Annual Meeting* (p. 15).

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. https://doi.org/10.1016/j.jml.2009.09.002

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181–218.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, *17*(3), 324–363.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237–288.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*(1), 69–83.

Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 843–853. https://doi.org/10.1037/a0029467

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*(1), 163–189.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377. https://doi.org/10.2307/2182440

Grice, H. P. (1975). Logic and conversation. *1975*, 41–58.

Grubert, A., & Eimer, M. (2016). Rapid attentional selection processes operate independently and in parallel for multiple targets. *Biological Psychology*, *121*, 99–108.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115.

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. https://doi.org/10.1016/S0749-596X(03)00022-6

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics*, *17*(4), 315–339.

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To Name or to Describe: Shared Knowledge Affects Referential Form. *Topics in Cognitive Science*, *4*(2), 290–305. https://doi.org/10.1111/j.1756-8765.2012.01182.x

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836. https://doi.org/10.1016/j.cognition.2008.04.008

Heyes, C. (2014). Submentalizing I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, *9*(2), 131–143.

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. https://doi.org/10.1080/01690960701402933

Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*(1), 1–35.

Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127–142. https://doi.org/10.1016/j.cognition.2004.07.001

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the Memory-Based Processing Approach to Common Ground. *Topics in Cognitive Science*, *8*(4), 780–795.

Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: The influence of speaker-specific memory associations on reference resolution. *Memory & Cognition*, *40*(1), 113–126.

Hübner, R., & Többel, L. (2012). Does attentional selectivity in the flanker task improve discretely or gradually? *Frontiers in Psychology*, *3*, 434.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

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., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, *39*(1), 1–20.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology*, *46*(6), 1417–1432. https://doi.org/10.1037/a0021062

Lagattuta, K. H., Sayfan, L., & Harvey, C. (2014). Beliefs about thought probability: Evidence for persistent errors in mindreading and links to executive control. *Child Development*, *85*(2), 659–674.

Lane, L. W., 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. https://doi.org/10.1111/j.1467-9280.2006.01697.x

Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1118.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556. https://doi.org/10.1016/j.jesp.2009.12.019

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512.

Lupianez, J., Ruz, M., Funes, M. J., & Milliken, B. (2007). The manifestation of attentional capture: facilitation or IOR depending on task demands. *Psychological Research*, *71*(1), 77–91.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*(2), 157–201.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. *Cambridge University Press*.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, *15*(5), 533–545.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *8*(3), 410–423.

Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two-and four-year-olds learn to adapt referring expressions to context: effects of distracters and feedback on referential communication. *Topics in Cognitive Science*, *4*(2), 184–210.

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), 749–760. https://doi.org/10.1037/a0019657

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*(3), 603–613.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, *28*(3), 292–312.

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., & Graham, S. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*(2), 220–249. https://doi.org/10.1016/j.cogpsych.2008.07.002

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition*, *148*, 136–144.

Piaget, J. (1959). *Judgment and reasoning in the child. Paterson, NJ: Littlefield, Adams, & Co.*

Piaget, J., & Inhelder, B. (1956). *The child's concept of space*. Routledge & Paul.

Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory*, *9*(4-6), 423–432.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515–526.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236. https://doi.org/10.1016/j.cognition.2010.08.003

Radvansky, G. A. (2012a). Across the event horizon. *Current Directions in Psychological Science*, *21*(4), 269–272.

Radvansky, G. A. (2012b). Across the Event Horizon. *Current Directions in Psychological Science*, *21*(4), 269–272. https://doi.org/10.1177/0963721412451274

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition*, *34*(5), 1150–1156.

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *The Quarterly Journal of Experimental Psychology*, *64*(8), 1632–1645.

Radvansky, G. A., Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: Environmental integration. *Psychonomic Bulletin & Review*, *17*(6), 900–904. https://doi.org/10.3758/PBR.17.6.900

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 608–620.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59.

Rich, S. S., & Taylor, H. A. (2000). Not all narrative shifts function equally. *Memory & Cognition*, *28*(7), 1257–1266.

Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. *Memory & Cognition*, *31*(8), 1284–1292.

Rubin, R. D., Brown-Schmidt, S., Duff, M. C., Tranel, D., & Cohen, N. J. (2011). How do I remember that I know you know that I know? *Psychological Science*, 0956797611418245.

Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-Taking in Comprehension, Production, and Memory: An Individual Differences Approach. *Journal of Experimental Psychology: General*. Retrieved from http://psycnet.apa.org/journals/xge/144/5/898/

Ryskin, R. A., Wang, R. F., & Brown-Schmidt, S. (2016). Listeners use speaker identity to access representations of spatial perspective during online language comprehension. *Cognition*, *147*, 75–84.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? *Methodology*.

Shintel, H., & Keysar, B. (2007). You said it before and you'll say it again: Expectations of consistency in communication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 357–369. https://doi.org/10.1037/0278-7393.33.2.357

Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, *1*(2), 260–273.

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science*, *10*(2), 199–204.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*(1), 125–140.

Sperber, D., & Wilson, D. (1986). *Relevance: communication and cognition*.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. In *Mind & Language* (Vol. 17, pp. 3–23).

Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and Automatic Perspective Taking in Children and Adults: Egocentrism and Automatic Perspective Taking. *Child Development*, no–no. https://doi.org/10.1111/j.1467-8624.2011.01730.x

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236.

Symeonidou, I., Dumontheil, I., Chow, W.-Y., & Breheny, R. (2016). Development of online use of theory of mind during adolescence: An eye-tracking study. *Journal of Experimental Child Psychology*, *149*, 81–97.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (1995). *Sentence comprehension. IN: J. MILLER & P. EIMAS. Eds. Handbook of Perception and Cognition Vol. 11: Speech and Language*. New York: Academic Press.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *J. Exp. Psychol*, *86*(2), 255–262.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154.

Twisk, J. W. (2006). *Applied multilevel analysis: a practical guide for medical researchers*. Cambridge University Press.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635–654.

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128–133. https://doi.org/10.1016/j.cognition.2013.10.004

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Citeseer.

Wang, J. J., Ali, M., Frisson, S., & Apperly, I. A. (2016). Language complexity modulates 8- and 10-year-olds' success at using their theory of mind abilities in a communication task. *Journal of Experimental Child Psychology*, *149*, 62–71.

Wang, J. J., Cane, J. E., Ferguson, H. J., Frisson, S., & Apperly, I. A. (submitted). *Why is it sometimes difficult to take account of a speaker's perspective?*

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review*, *20*(4), 766–772. https://doi.org/10.3758/s13423-013-0396-1

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 655–684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Wu, S., & Keysar, B. (2007). The Effect of Culture on Perspective Taking. *Psychological Science*, *18*(7), 600–606. https://doi.org/10.1111/j.1467-9280.2007.01946.x

Yoon, S. O., Benjamin, A. S., & Brown-Schmidt, S. (2016). The historical context in conversation: Lexical differentiation and memory for the discourse history. *Cognition*, *154*, 102–117.

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 919–937. https://doi.org/10.1037/a0036161

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, *19*(4), 699–707.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84.

Zhao, L., Wang, J. J., & Apperly, I. A. (under review). *Remembering what she knows: The costs of holding in mind a speaker's perspective during a communication task.*

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 292–297.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 386.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, *2*(3), 199–220.