Perspective Taking in Definite Reference Resolution

Xiaobei Zheng

Thesis submitted for the fulfilment of the requirement for the degree of

Doctor of Philosophy in Linguistics

UCL

2015

Declaration:

I, [Xiaobei Zheng] confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:___________________ Date:__________________

Acknowledgements:

I am using this opportunity to express my sincere gratitude to my supervisor Dr Richard Breheny for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I am honoured to have been his PhD student, and to have been guided, challenged and inspired by him.

I especially express my warm thanks to my colleague Irini Symeonidou, Giulio Dulcinati, and Chao Sun for testing participants for my study when I was in China. I am thankful to Ye Tian, Chao Sun, and Irini Symeonidou for sharing their truthful and illuminating views on a number of issues related to the experiments. I also would like to thank my second supervisor Dr. Nausicaa Pouscoulous for her advice and encouragement when I start my thesis writing.

Throughout my PhD journey, I have received unfailing assistance and guidance from many people at the Department of Linguistics, UCL, including Robyn Carston, Wing Yee Chow, Nathan Klinedinst, Ad Neelman, Andrew Nevins, Matthew Reeve, Andrea Santi, and Hans van de Koot. I also much appreciate the discussion, support and the friendship from the people in PhD room, including Zoë Belk, Harris Constantinou, Natalia Cichosz, Thiago Galery, Felicity Deamer, Giulio Dulcinati, Elizabeth Eden, Matthew Gotham, Claire Grant, Sam Green, Jiri Kaspar, Akis Kechagias, Ya-Fang Lu, Diana Mazzarella, Nick Neasom, Isabelle Needham-Didsbury, Thanasis Soultatis, Chao Sun, Irini Symeonidou, Kevin Tang, Ye Tian, and Maria Varkanitsa.

Last but not the least, I would like to thank my family: my parents, my husband, and my daughter for supporting me spiritually throughout writing this thesis and my life in general.

Abstract

This thesis explores the roles of common ground and Theory of Mind in processes for definite reference resolution. It examines the proposals due to Herb Clark and colleagues that common ground is common-knowledge of discourse relevant facts and that it serves as the main constraint on definite referential processes. The thesis contrasts that view with an alternative where common ground emerges from shared interactions as a result of low-level, or automatic memory mechanisms that are constrained by what has been in shared attention. The thesis presents two experiments that verify the effect of memory-based mechanism on referential processes, and also shows that memory-based mechanisms function only where the cue to previous shared experience is itself in shared attention.

Given that common ground only weakly constrains definite referential processes, it is argued that Theory of Mind inferences play a critical role in on-line referential processes where interlocutors have different perspective information. The thesis argues that research using the perspective-taking task did not distinguish between the role of common ground and inferences about the speaker's ignorance of certain objects. Experiment 3 shows that participants have better perspective taking when the procedure highlighted the speaker's ignorance. This effect is seen within 300ms of the on-set of the critical linguistic stimulus, demonstrating very early integration of Theory of Mind inferences. Then a follow-up experiment excludes the possibility of the effect of simple interactivity in the experiment 3, and also explores whether the co-presence of the speaker increased the degree to which the existence of private objects was shared attention. Because the existence of privately viewable objects are common ground, even though their specific identities are not. The results of experiment 4 support the proposals about common ground in this thesis by showing that participants have better performance when the speaker is not co-present.

Keywords: common ground, perspective taking, definite reference, Theory of Mind, .

Table of contents

List of Tables

List of Figures

Introduction: Social cognition and definite reference

Cognitive processes for successful definite reference must integrate information from a wide number of sources. These sources include linguistic knowledge, the utterance situation and general world knowledge. In addition, referential processes rely on information about our interlocutors as social agents since language use is a highly social activity. In both the theoretical literature and the psycholinguistic literature, it has been proposed that two constructs play a key role in decision processes for definite reference. These are common ground and Theory of Mind. The idea of Theory of Mind comes mostly from developmental and ethological research (Premack et. al., 1978; Leslie, 1987; Leslie et. al., 1988). It refers to a set of abilities that span a wide range of domains of use and involve making inferences about agents' behaviour based on causally active, unseen mental states. In particular, these states should include some correlate of belief, desire and intention (see Lewis 1994 for a sketch of such explanations). The idea of common ground comes from the philosophical literature and linguistics literature. Broadly speaking, common ground is thought of as a repository of information that interlocutors 'share', in some sense, and take for granted as shared. As will be discussed in the coming sections common ground is normally tied quite closely to the idea of 'common knowledge' (or 'mutual knowledge'). Common knowledge is a theoretical construct that is used to account for how commonly held information could be 'open'. Common knowledge of some fact between two agents requires those agents to be able to represent the other agent as capable of holding beliefs or other attitudes to states of affairs. Thus, a common assumption in the literature is that processes for representing common ground are tied into Theory of Mind abilities. In widely discussed psycholinguistic research due to Herb Clark and colleagues on how reference is made with definite noun phrases, all of these constructs come together: Common ground is common knowledge of discourse relevant facts, definite referring expressions are constrained to refer to objects in common ground.

In this thesis, I will argue for the following position:

i. that common knowledge plays no role in processes for definite reference;

ii. that common ground is a product of social-cognitive processes that are not linked into full-blown theory of mind abilities;

iii. common ground plays a role in definite referential processes only by promoting attention to previously shared referents;

iv. where perspective-taking is required in referential decision processes, inferences based on Theory of Mind abilities are necessary for success, common ground plays a peripheral role (as per (iii) above).

In Chapter 1, I will review the theoretical literature on definite reference and common ground. There we will critically evaluate the proposal, due to Herb Clark, as well as some philosophers and linguists, that common ground is common knowledge of discourse relevant facts and that definite referents are strictly constrained to be commonly known. I will argue that this position has little support when seen as a proposal about actual cognitive processes. Instead, I will follow the majority view in linguistics and philosophy that the uniqueness of definite reference is key.

I will then review research suggesting that mechanisms for tracking shared information (common ground) among agents develops in infancy, independently of full-blow Theory of Mind. Turning to consider proposals about tracking shared information (or common ground), we will see that an important set of recent research sees domain-general mechanisms for memory and priming as playing a key role. We will consider this as a potential basis for maintaining common ground cognitively, but argue that the mechanisms are constrained by social-cognitive factors.

In Chapter 2, I will present a set of studies that test the extent to which memory-based mechanisms for common ground are domain general. These studies will demonstrate that memory-based mechanisms are constrained at the coding phase to information that was previously in shared attention.

In Chapter 3, I turn to consider the considerable body of research on when and how participants take their interlocutor's perspective in on-line referential decision processes. Specifically, I will focus on the referential communication game, or perspective-taking task where what participants and their interlocutors can see is different. One aim of this chapter is to discuss a widely adopted model of how participants might successfully determine the correct referent in these tasks, based on the ideas of Herb Clark and colleagues. I observe that this model implies that success on perspective-taking tasks could be achieved *without* direct involvement of Theory of Mind inferences.

A second aim of this chapter is to set out what are the respective roles of Theory of Mind and common ground in perspective-taking tasks. It will be proposed that in the most widely used perspective-taking tasks, common ground mechanisms in fact may work against processes for adjusting to speaker's perspective.

In Chapter 4 I present a series of studies that seeks to disentangle the respective roles of Theory of Mind and common ground in perspective taking tasks. It will be shown that cues that promote theory of mind inferences play a far more potent role than those for common ground, although the latter are a factor in such processes.

Chapter 1: Common ground as the target of shared attention

1. A role for common knowledge, or mindreading in definite reference?

1.1 Two theories of definite reference, common knowledge and common ground

Consider a scenario where two people are shopping in the supermarket. When they walk past a cheese refrigerator, one person says to another “please hand me the cheese”. The addressee will be confused about which specific cheese the speaker wants. This utterance is less effective in that the speaker does not provide a definite description that is as specific as is necessary in order for the interlocutor to identify the intended referent (Russell, 1919; Neale, 1990). However, if the speaker has just been talking about a recipe using Mozzarella, the addressee will understand the utterance “the cheese” as, ‘the mozzarella in the cheese fridge’. The interlocutor presumes that their partner is a cooperative agent and is able to keep their conversation neither over-informative nor under-informative (Grice, 1975). In order to communicate an intended referent successfully, the speaker and the addressee know that the indicated entity is part of a relevant message in the broader context that they share, and presumes each other are able to make inference about other's mental states quickly (Sperber & Wilson, 1986).

In this case of definite reference resolution, what contributes to the addressee searching for the indicated referent within a given domain? It seems clear that in conversation, people have to make inferences about what is known by the other interlocutor. Many have claimed that we not only rely on our interlocutor being able to make a decision about a definite reference based on what we assume the other knows, or even what the other knows about what we know but about what we in a sense, 'know together' (Tomasello et al. , 2005).

According to a widely adopted theory, both the speaker and the addressee constrain the domain for definite reference to what is common knowledge or mutual knowledge (Clark & Marshall, 1981; Schiffer, 1972). In a related approach to definite reference and presupposition, it is thought that the referent of a definite description is 'given' or presupposed in what is held to be common ground (Jespersen, 1949; Heim, 1982; Clark, 1992). Stalnaker (1974, 1975, and 2002) defines common ground in terms of common knowledge of discourse relevant facts[1]. Dynamic approaches to definite reference and presupposition (Heim, 1982, 1983) adapt Stalnaker's ideas to model how utterances update a shared information state (the dynamic context) which consists of what is taken to be commonly known or assumed. Thus common ground is typically analysed in terms of common knowledge (see Stalnaker 2002 for an explicit formalisation of this). Where this assumption is made, *common ground* and *common knowledge* (of discourse relevant facts) are interchangeable ideas.

Common knowledge of some fact is not just a matter of whether interlocutors known they both know that fact. Common knowledge is typically illustrated as follows (using the above example): we express proposition that *there is mozzarella cheese available in the cheese fridge* as *p*. First people A and B share knowledge of *p*:

1) A knows *p*.

1') B knows *p*.

Second people A and B know they share the knowledge of p

[1] In fact, Stalnaker's discussion of common ground is more carefully cast in terms of what interocutors are willing to take for granted as commonly known, for the sake of the conversation. But the structure of the commonly held assumptions remains the same in terms of Stalnaker's analysis and can be traced back to proposals in Lewis (1969).

2) A knows that B knows p.

2') B knows that A knows p.

Finally, according to certain formalisations of the idea of common knowledge due to Schiffer (1972), Clark & Marshall, (1978); Aumann (1976) and others, people A and B mutually knowing about p implies the following infinite sequence:

3) A knows that B knows that A knows about p.

3') B knows that A knows that B knows about p.

et cetera ad infinitum.

Formalisations of common or mutual knowledge that have these infinite implications at best apply to ideally rational agents with unlimited capacity to reason and so have been criticised as being cognitively implausible (Skyrms, 1996; Sperber & Wilson1986; Cubitt & Sugden 2003; Paternotte 2011). Paternotte (2011) makes the point that even ideally rational agents with normal finite capacities for perception, memory and inference would not be able to establish common ground in this way, due to the fallibility that finite cognitive resources entail. That is, if one knows that an agent's abilities of perception, memory and inference are limited and liable to the occasional error, then each level of meta-representation in (3) involves more risk that the relevant facts have not been attended to, the relevant inferences have not been made, involving the relevant information from memory, and so forth.

Lewis (1969) and Sperber & Wilson (1986) provide more psychologically plausible accounts of common knowledge that do not imply this infinite series of higher-order beliefs but do allow interlocutors to infer a limited series of higher-order beliefs where necessary and where other contextual information about shared mental states allow (see Cubitt & Sugden, 2003; Paternotte 2011 for a lengthy discussion). These accounts of common knowledge (or mutual manifestness) are cast in terms of *potential* attitudes to situations. Lewis defines common knowledge in terms of a notion of 'reason to believe' rather than belief. In Lewis' system, it is possible that A and B have common knowledge of p without either having any actual beliefs about p at all (Lewis 1969, p.55). This is because some external state of affairs can provide two agents with reason to believe that the external situation gives both agents reason to believe some fact (that becomes, thereby, mutually known). Such situations Lewis calls, 'reflexive common indicators'. Lewis, (1969) gives an example of a reflexive common indicator: in a town where the telephone system is unreliable, people normally wait for the original caller to call back if the current call is interrupted. These past events become salient when a call is interrupted and their existence suffices to be a reflexive common indicator for common knowledge of the convention that caller calls back in that town.

Another important reflexive common indicator to be discussed at length below involves the visible signs of shared-attentional activity. These are signs of triadic attentional behaviour between two agents and the object of shared attention. These signs provide evidence that the object of attention's existence is in shared attention, and hence common knowledge (Barwise, 1988/1989; Peacocke, 2005). An important point to note here is that it may be necessary to rely on Level-1 perspective-taking abilities (Flavell et al., 1981) in order to recognise that the situational, or behavioural signs of shared attention are present. As widely discussed in the developmental literature, such visual perspective taking abilities should be considered distinct from Theory-of-Mind abilities, as they are conceived of in this thesis (see Baron-Cohen et al., 1985).[2]

[2] A terminological note. The label, 'Perspective-Taking Task' seems to derive from a usage that began with Baron-Cohen et al. (1985) – see Brown-Schmidt & Hanna (2011). Baron-Cohen et al. counterpose visual perspective-taking (as studied in Flavell et al. 1981) with what they call, 'conceptual perspective-taking'. The latter, Baron-Cohen et al. argue, is involved in false-belief

Thus, according to these proposals, the existence of common knowledge is more about external factors, coupled with human cognitive abilities, that provide an evidential basis for these potential attitudes. And so, according to Lewis' original proposals about common knowledge, in order to confirm that they have common knowledge of p, two people do not need to check that an infinite series of ever more complex representations of mental representations hold, there simply needs to an evidential basis for common knowledge, a 'reflexive common indicator'. Clark & Marshall (1978; 1981) discuss a number of environmental cues that could provide such a basis: (i) joint attention to a co-present object, (ii) mention in previous discourse, (iii) shared cultural or background knowledge.

Thus, an account of definite reference that requires the target to be common knowledge in the sense defined by Lewis or Sperber & Wilson may be cognitively plausible. Since it does not require interacting agents to have an infinite series of beliefs about other beliefs, it only requires there to be enough evidence that there is a reflexive common indicator to common knowledge.

However, not all approaches to pragmatics or definite reference endorse the idea that an object needs to be in common ground in order for definite reference to proceed. In fact, in the long history of research into definite reference, perhaps the most famous and widely adopted proposal, due to Bertrand Russell (see Russell, 1905, 1919; Neale, 1990) specifically denies the idea that definiteness is directly related to givenness. Rather, Russell's proposals about definites are that there is a unique object in the relevant context that satisfies the description.

---

tasks, such as Sally-Anne. They also argue that the description of false-belief tasks should be framed in terms of Theory of Mind. Visual perspective-taking concerns a separate ability: to grasp that visual perception relies on the spatial arrangements of objects and the functioning of the eyes.

In this thesis, I use 'Perspective Taking Task' to refer simply to the experimental tasks of the kind reported in the psycholinguistic literature reviewed. In line with Baron-Cohen et al., I sometimes use 'speaker's perspective' to refer to their relevant mental states. Whenever visual perspective-taking is at issue, I use, 'visual perspective-taking'.

To sum up this discussion, there are two views on definite reference. According to Clark & Marshall (1981), and others, definite expressions obey the common ground constraint:

Common Ground Constraint: A definite expression is felicitous only if its referent is common ground when it is used.

According to others, that follow Russell (see Neale, 1990), the denotation of a definite expression has to be unique. This leads to an identifiability constraint:

Identifiability Constraint: The referent of a definite description is identifiable for the hearer, under the given description.

In this thesis, the aim is not to adjudicate on whether Russell is right or whether some version of the givenness hypothesis is correct. Our aim is to establish the cognitive mechanisms for establishing the referent of a definite expression in on-line processing. In the rest of this section, I consider Clark's motivation for his version of the common ground constraint, involving common knowledge of the target referent. I will also consider the phenomenon of so-called accommodation, and what that means for the use of common knowledge in on-line definite processes.

1.2 The common knowledge constraint and accommodation – implications for models of referential processes.

In the tradition of Russell, Grice (1957, 1982) remained unconvinced that common knowledge was necessary at all in communication, pointing out that for practical purposes, all that people are liable to make inferences about are what the other interlocutor may know or not know, what the other person knows about what oneself may know or not know and little more.

Perhaps surprisingly, this sentiment is echoed by Lewis himself (Lewis, 1969: p.32). That is, Lewis recognised that for practical purposes, the warrant for a co-ordination decision would rarely involve more than a few iterations of what the other expects about what the self expects, what Lewis calls, 'higher-order expectations'.

So, if Lewis assumes that two agents are only likely to make inferences at relatively few levels of metarepresentation, why does he include common knowledge in his analysis of games of co-ordination? [3] The answer is simply that he includes common knowledge only because he could not think of a case of such activity that did not seem to involve something like common knowledge (ibid: p. 59). In other words, Lewis does not argue that common knowledge features in social encounters involving co-ordination because it is *necessary* for common knowledge to be present. The claim is that it happens to be an ever-present feature of such encounters. In a similar vein, Sperber & Wilson (1986) include their version of common knowledge (mutual manifestness) into their definition of communication only because intuition suggests that communication involves a specific kind of 'overtness' or, 'openness' of the speaker's informative intention that their notion of mutual manifestness is designed to capture. Thus, they propose that in communication, the speaker intends that the intention to inform is mutually manifest, but that is all. Regarding the intention to inform (and the intention to refer), this is not made mutually manifest but merely manifest.[4] Lewis and Sperber and Wilson are not alone in this kind of view of 'common knowledge'. Since the work of Grice, philosophers and psychologists report the intuition that acts of communication (Schiffer, 1972, Harman, 1977), and joint activities (Bratman, 1992; Tomasello et. al., 2005) are always accompanied by a

[3] The topic of Lewis (1969) is convention, which Lewis analyses in terms of games of co-ordination. As Lewis explains, many kinds of social encounters pose problems of co-ordination among the agents involved. In communication using definite reference, there is a problem of co-ordination which is just the problem of the speaker choosing a definite form with an intended referent and the hearer inferring that the form has the same referent that the speaker intends.

[4] So, like Grice, Sperber and Wilson assume that when it comes to co-ordinating on definite reference, one takes higher-order thoughts about other's thoughts into account only where necessary.

special type of ‘open awareness’, ‘overtness’ or ‘knowing-together’ that the Lewis’ theoretical concept of common knowledge, or variations on that have been used to describe.[5] I will consider why common knowledge or, ‘overtness’ seems to be an ever-present feature of communicative acts in later sections.

Lewis’ point of view stands in stark contrast to that put forward in Clark & Marshall (1981). There the argument is that common knowledge of the relevant facts for establishing a definite referent is necessary since otherwise interlocutors could not be certain that they are thinking about the same referent. Clark & Marshall argue that certainty about the referent is only guaranteed by an infinite sequence of higher order assumptions about the relevant facts and that it is natural and rational to require such certainty in conversation. But as mentioned above (see Paternotte, 2011) even ideally rational agents would not consider such certainty a realistic goal due to the known finite limitations of human perception, inference and memory. It is interesting that Clark & Marshall (1981) follow Lewis in proposing that environmental cues (such as co-presence) are used as the evidential basis of common knowledge of some fact. Thus Clark & Marshall (1981) proposals about how common knowledge is established and maintained are cognitively realistic. However, their motivation for why common knowledge should be a feature of definite references rests on assumptions that are not.

So, if we have an account of common knowledge that makes it plausible that interlocutor’s keep track of relevant commonly known information, does that not mean that The Common Ground Constraint is plausible? As we are about to see, the phenomenon of accommodation means that the Common Ground Constraint cannot plausibly guide referential processes, it only describes their outcome. By contrast, we will

[5] Other notable early attempts to define common (or ‘mutual’) knowledge are found in Schiffer (1972) and Aumann (1976).

see that Theory of Mind abilities are often required to be in play when psychological mechanisms for maintaining common ground could not.

Suppose that Sue buys some cheddar cheese, brings it home and puts it in the fridge next to some mozzarella that is already there. Later her roommate Jane makes a dish. She asks Sue, ‘Can you pass me the cheese from the fridge’. Looking in the fridge, Sue sees two kinds of cheese. Since Sue was not there when Jane put mozzarella in the fridge and Jane was not there when Sue put the cheddar in the fridge, we can say that it is not common knowledge that either cheese is in the fridge. But Sue can infer that when Jane put mozzarella in the fridge, there was no other cheese in the fridge. So, Sue can see that from Jane’s perspective, ‘the cheese’ is felicitous since there is only one cheese in the fridge. These inferences, do not rely on what is common knowledge (since nothing is actually common knowledge) but they do rely on Sue making inferences about Jane’s mental state.

Examples similar to this kind have been widely discussed in the theoretical literature on common ground and presupposition where it is proposed that there exists a mechanism of ‘accommodation’ that allows one to retrospectively introduce an object into common ground upon the use of a definite description (Stalnaker, 1974; 1978), and hence become common knowledge. So the proposal is that Jane expected Sue to retrospectively accommodate the mozzarella cheese as part of common ground when she hears, ‘the cheese’. While it is certainly clear that the mozzarella cheese becomes part of common ground (or common knowledge) once the utterance is made and understood, from a *cognitive* point of view, it could be argued that any mechanisms employed for maintaining common ground or common knowledge play no real role in establishing the referent of the noun phrase, since inferential work to establish the referent of the definite happens before the accommodation of the referent into common ground. At the same time, inferences about the speaker’s mental state (what she is likely to believe) including her referential intentions (how the speaker is likely to expect the hearer to respond to her utterance) do play a critical role.

As a rule, it could be argued that where accommodation of a definite referent occurs (as it often does), the cognitive process required to establish the referent does not call on processes involved in maintaining common ground (since the target is not in common ground). Of course can still accept that there are mechanisms for maintaining common ground and that common ground objects do figure prominently in referential processing (for both definite and indefinite descriptions). The point is only that, from a cognitive perspective at least, it is not necessary that an object be actually represented as common ground (or to be commonly known) to be a definite referent.

There is however a further point about when common ground may play a role in referential processes. Consider an alternative scenario to the one above where both Jane and Sue put both the cheddar and mozzarella in the fridge together. Here it is common ground that both cheeses are in the fridge. Still, when making a dish that is known by both to require mozzarella cheese, it seems Jane can felicitously ask Sue to, ‘pass the cheese from the fridge’. Here, although it is common ground that there is no unique cheese in the fridge, there is just one cheese that is relevant to Jane’s current purposes. Thus, processes that determine how the utterance is relevant seem to further select from common ground. It seems then that, from a cognitive perspective, common ground is neither necessary nor sufficient for definite reference.

The first example discussed above involved a case of so-called accommodation. In that case, what is accommodated is the existence of a specific referent for a definite description. Here the referent is identified by the hearer, Sue, via visual perception. Among Clark & Marshall’s categories of common knowledge, we also find common knowledge of some fact due to membership of some community or culture. For example, common knowledge of the current US President. We can consider another example of accommodation in definite reference which involves such a ‘cultural fact’. Consider now a case where it is common knowledge that Jane and Sue placed two cheeses (mozzarella and cheddar) in the fridge. Jane is preparing a dish that Sue is unfamiliar with but it seems to be according to an Italian recipe. It is not common ground between Jane and Sue that the recipe calls for mozzarella cheese. However, when Jane asks Sue to pass the cheese from the fridge, Sue can reason that the kind of dish Jane is preparing is much more likely to call for mozzarella than cheddar and she can see that Jane would expect her to see that. Thus, Sue reasons that the intended referent is the mozzarella cheese. In this case a piece of ‘background’ information necessary for establishing a referent is not in common ground (not commonly known) when the utterance is made. Inferences about Jane’s likely intentions and expectations are required to establish the relevant background information, but these inferences call on information that is not common ground. Once communication succeeds, however, the relevant piece of information (that the dish Jane is preparing calls for mozzarella) becomes ‘accommodated’ into common ground.

In this section, I have argued against the idea that the Common Ground Constraint could apply in on-line processes for definite reference. That is, viewed in terms of cognitive processes, common ground or common knowledge would not be essential to referential processing. At best, it would serve only to provide a record of previous shared experience in conversation, once the referential and communicative intentions are established. Thus there are grounds, based on merely introspective evidence to discount common ground or common knowledge as being a factor in processes for definite reference. This is not to say, however, that something like common ground, at least, has a real presence in social and communicative interactions. There is reason to think that ‘common ground’ may be underpinned by distinctive cognitive mechanism that develop in infancy and underpin social interactions generally. We turn to review this evidence next.

2. The 'openness' of social interactions and shared attention

In the last section, we considered theoretical accounts of common ground that attempt to analyse the notion in terms of some form of common knowledge, i.e. some form of (potential) mental state that is about the (potential) mental states of oneself and others. Thus, from a psychological perspective, common knowledge and common ground could be viewed as constructs relying on human abilities to represent and reason about others' mental states.[6] This ability is often referred to as Theory of Mind (ToM) ability (Leslie, 1987). It was also mentioned that leading research on the place of common knowledge in social interactions diverges on the motivation for its presence. On the one hand, Clark & Marshall (1981) motivate a need for common knowledge on the basis of an argument about certainty: speaker and hearer cannot be certain they are thinking about the same referent unless it is common knowledge. But, as suggested in the last section, this line of thinking is flawed: Interlocutors can only attain certainty about co-ordination if common knowledge is defined along the lines of Aumann (1976), implying an infinite series of ever higher-order beliefs about the other's higher-order beliefs. But it is not only not possible for cognitively finite individuals to attain such a state, an ideally rational, but cognitively limited agent would understand that it is incoherent to expect the parties involved to attain such a state. Thus absolute certainty about referents is not a goal that rational agents should expect.[7] This is not to say that if agents have common knowledge, as Lewis defines it, they cannot adduce further support for their decision about a referent on the

[6] Even if, as Lewis proposes, common knowledge of p by two agents does not require those agents to even represent p, it requires the agents to be able to represent the other's beliefs and beliefs about beliefs etc.

[7] Note, this line of thought is consistent both with general skepticism about knowledge and with an acceptance that knowledge is possible. The argument is that even though knowledge by an individual may be possible, ideal mutual knowledge is not.

basis of that common knowledge. As Lewis (1969 pp.28-32) demonstrates, each higher-order expectation of the other agent's higher-order expectation can provide a little further support for one's decision in a co-ordination problem. But as we saw in the last section, it is neither necessary nor sufficient for the referent of a definite description to be common knowledge and in fact, interlocutors rarely go beyond second- or third-order expectations in the process of seeking evidential support for a decision.

In contrast to Clark & Marshall, Lewis (1969), Sperber & Wilson (1986) and many others motivate the presence of the 'overtness' or 'openness' involved in co-ordination in social interaction based on introspective evidence alone. For communication, the idea is that cases where the communicator has an informative intention that is not in some way, 'public' or 'overt', are not intuitively felt to be cases of 'communication' in one key sense of that word. Assuming these philosophers' and linguists' intuitions are on the right track, the question is raised why we have this sense about human communication.

One important part of the answer to this question comes from infant research. Csibra (2010) reviews a number of studies of infants between 0-12months. These studies reveal an acute sensitivity to 'ostensive cues', such as eye contact, contingent behaviour (as found in turn-taking), and self-directed speech.[8] For instance, a series of studies by Farroni and colleagues (Farroni et al., 2002, 2004) show that 4-5month old infants respond to shifts in eye-gaze only if they are engaged by direct eye contact prior to the shift of attention. Similarly, Johnson (2003) and Deligianni et al (2011) show that children respond to an agent's ambiguous gesture as being communicative only when the agent engages in contingent, rather than independent, behaviour with the infant prior to the

[8] Csibra talks in terms of 'infant-directed speech' as being an ostensive signal. However, it is not clear whether it is the distinctive prosodic properties of IDS that constitute the relevant cue or the fact that IDS is understood by an infant as speech directed at the self. Csibra adds to the list of cues for slightly older children the child's name being uttered. Thus it could be it is speech that the child discerns as being directed to itself that is the cue, rather than any specific prosodic properties of IDS and that the child simply associates IDS prosody with self-directed speech.

gesture. Grossman et al., (2008) present a NIRS study revealing that regions of the infant brain activated by eye-contact correspond to key regions involved in processing communicative gestures. Based on such evidence, Csibra (2010) conjectures that infants have an innate pre-disposition or bias to process stimuli such as eye-contact as ‘ostensive signals’. These would be signals that infants simply decode as meaning that the agent of the signal has an informative intention, i.e. wants to convey some information. In Csibra & Gergely (2009), it is proposed that infants are biased to expect a specific type of relevant information by these ostensive signals. Thus for Csibra and colleagues, cues such as eye-contact or contingent behaviour stimulate an expectation in infants that they will receive relevant information.

It is possible to question some of the details of Csibra’s proposal. In particular, it is an open question whether all of these stimuli are solely signals that the agent has a communicative intention. It could be, for instance, that contingent behaviour is a cue for shared interactions in general. It seems reasonable to wonder whether turn-taking could occur without any attendant communication, in for instance so-called joint action. This is a question of some interest in the literature on joint action and shared attention (see Carpenter & Liebal 2011). An alternative interpretation of the results reviewed by Csibra would be to say that these are cues to something like what Tomasello et al. (2005) call, ‘shared intentionality’. This is a state that applies to all kinds of interaction, regardless of whether it is specifically communicative. Tomasello et al. (2005) note that it is a feature of all such states that they are open in the same way, involving a feature termed, ‘knowing together’ (Tomasello’s version of common knowledge).

But putting aside questions about the details of the theoretical proposal, it seems clear that infants do process stimuli differently depending on whether the stimuli are accompanied by these specific cues and that in many cases infants seem to treat the actions of the carer as communicative. These same cues, or at least eye contact and contingent reactivity are present in what is termed joint attention in infants. Joint attention is among the earliest forms of 'triadic' behaviour where adult and infant share experience of a third object or event. It has been argued that joint attention is a critical pre-curser to common knowledge (Barwise, 1988/1989; Peacocke, 2005) having the self-referential properties required to allow agents to infer as many levels of higher-order expectations as they need. In Barwise's terms, shared attention among A and B to some object, x, involves A and B attending to the situation in which they both attend to x. This kind of 'shared situation' fulfils Lewis's definition of a 'reflexive common indicator' that x is co-present and is, in Lewis' theory, a basis for common knowledge of x's co-presence.

Note we assume, following Carpenter & Liebal (2011), Peacocke, (2005) among others, that attention in shared attention is an intentional activity. This suggests that to recognise an interaction as shared attention, one needs an operating Theory of Mind. However, there is a robust body of research pointing to the conclusion that infants engage in joint attentional activity long before their first birthday. We follow Peacocke and Carpenter & Liebal who accept this but assume that joint attentional activity engages a limited amount of Theory of Mind in very young infants. We believe that this is possible due to the fact that infants are disposed (perhaps innately) to respond to cues to shared attention. Csibra (2010) surveys a still-growing body of research showing that infants respond to eye contact, contingent reactivity and self-directed speech in a manner that suggests a bias to process the subsequent object of attention more deeply for relevance. In Barwise's terms (see Barwise, 1988, 1989) when we share attention to an object situation, *s*, we both attend to *s* and attend to the broader situation in which we attend to *s*. Shared attention yields what Barwise terms a 'shared situation'. It therefore seems to follow from infant social-cognitive research that humans have a specific receptivity to shared attentional cues that manifests itself in processing biases concerning the shared-situational source. It is useful here to draw a distinction between having a concept and having a conception (see Fodor, 1998). To have a concept of x is to be able to reliably discriminate x. To have a certain conception of x is to be able to draw specific inferences about x. A child may have a concept of a car without any conception of the functioning of the internal combustion engine. Similarly, a child may have a concept of shared attention, be able to apply that concept in recognition of certain situations, but may not have a conception of its intentional nature until after 12 months of age.[9]

Thus in infant experience, cues to communication and potentially to other joint activities that require co-ordination are bound up with a specific joint activity that can explain our sense of openness or publicness cited by philosophers as being constitutive of communication and joint action. It seems a small step then to conjecture that our sense of what communication is stems from our developmental experience. Communication being an activity that not only results in a form of shared or joint attention to what is being communicated but seems to require joint attention to the act of communication itself.

3. The effects of joint engagement on subsequent communicative processes

If we consider research on the results of shared activities in slightly older infants, from their first birthday onwards, we can get a sense of the cognitive processes that are triggered by stimuli such as eye-contact and contingent behaviour. An important study demonstrating toddlers' appreciation of what is shared and what is not is reported in Tomasello & Haberl (2003). Here 12-month-old infants share attention to two novel objects with an experimenter. In the subsequent absence of the

[9] Note that in Section 1.1 above, I clarified that even though recognising the situational signs of shared attention may require Level-1 perspective-taking, this ability is distinct from Theory of Mind, as that is conceived in this thesis.

experimenter, the infant is given a third novel object with which to engage. All three objects are then placed together. In the test phase, the experimenter returns and gestures excitedly but ambiguously toward the three objects using only the singular demonstrative, ‘that’ to refer to an object. Tomasello & Haberl report that infants select the third, unshared object for the experimenter. This is interpreted as evidence that infants appreciate what might be novel for the experimenter. In a follow-up study, Moll & Tomasello (2007) used three groups of children in separate conditions. A shared interaction condition replicated the procedure of Tomasello & Haberl (2003), having the child and experimenter jointly interact with the first two objects. An individual engagement condition saw the child and the experimenter separately engage with each of the first two objects, although in each other’s presence. A third, on-looker condition saw the child engage with the first two objects alone while the experimenter was simply an on-looker. At test, it was found that for younger infants (14 months old), only the joint engagement resulted in choosing the third object. For older, 18 month-old, infants also individual engagement resulted in the choice of the correct object, but not merely the presence of an on-looking adult. These results suggest that prior joint engagement plays a key role in allowing infants to take perspective. It is not clear from Moll & Tomasello (2007) however whether the perspective infants use is that of the experimenter them self or the previous ‘joint perspective’.

That infants use individual agents as a cue to previous joint perspective is established in Liebal et al (2010). In this study, an infant would interact consecutively with two experimenters in two separate activities. At test, one of the experimenters would gesture toward an object that could be involved in either of the activities that the child has been involved in. Liebal et al. report that 14- and 18-month old infants interpret the gesture toward the object relative to the previous joint engagement with the experimenter that makes the gesture. This is even the case when the relevant activity was not the most recent activity that the child engaged in. The interpretation of these results is that infants use previous shared experience as the source of relevance for the interpretation of a communicative gesture. This is the case even where that previous shared experience happens prior to a more recent, potentially relevant experience. The results suggest that the experimenter themselves acts as a very strong cue to re-activate the memory of the previous shared engagement, this is a phenomenon that is explored also in adult processing literature, to which we will return below.

The results from Liebal et al (2010) and similar results reported in Southgate et al (2009) point to a mechanism whereby other agents can provide a strong cue to the reactivation of previous shared experience. A second follow-up to Tomasello & Haberl (2003) reported in Moll et al. (2011) tests the idea that lower-level mechanisms may be an important causal factor in the effect of previous joint engagement in providing relevant context for subsequent joint engagement. This study again uses a design based on Tomasello & Haberl where experimenter engages with the child interactively to explore two novel objects. In a second phase, the child is given a third novel object that the experimenter cannot see. In one condition, replicating Tomasello & Haberl's procedure, the experimenter leaves the room during this second phase. In three other conditions, the experimenter remains present (though visual access to the third object is excluded) and/or engages with the child verbally whilst visually absent. Moll et al. found that two-year old children chose the correct item at a rate above chance in the silent absence condition (replicating Tomasello & Haberl 2003). In the other conditions, children were not above chance, while rates of correct choice were significantly below those of the baseline condition in both of the presence condition (independently of whether the experimenter engaged with the child verbally).

There are implications of this study for the question raised by the original results in Tomasello & Haberl (2003) – whether children succeed by considering the experimenter's individual perspective, or they considered the gesture from the previous shared perspective. It seems that the co-presence conditions in Moll et al.'s study provide sufficient cues to the child that their experience of the third object is shared with the experimenter, even though the child is made aware that the experimenter cannot see the third toy. This suggests that younger children succeed at this task by utilising previous shared perspective, as supported by lower level cues. That is, it would be a shared perspective that does not integrate information derivable through Theory of Mind reasoning (for otherwise the child knows that the third toy is novel for the experimenter). Thus something like shared perspective could be maintained via a lower-level system that relies on cues that come with interactions without necessarily integrating information that could be inferred using one's theory of mind. It squares with the fact that infants throughout the first year of development rely on these cues in order to share experience and to perhaps provide access to relevant past experience.

4. The cognitive foundations of common ground

Csibra (2010) speculates that there is an innate set of biases in infant cognition that respond to certain cues (his ostensive signals) by devoting attentional resources to process information in the shared situation as highly relevant. Thus, any subsequent stimulus involving the agent that child has interacted with and that past shared situation is liable to become re-activated to a greater extent than information that may not have been previously attended to as shared, since unshared stimuli do not receive as much attention. Thus, if Csibra's conjecture is correct, there is a cognitive basis for maintaining and prioritising a record of previous shared experience. Below, we will review research on memory-based mechanisms involved in language processing. We will propose that the cognitive biases found even in infancy modulate the functioning of memory mechanisms that make previously shared information accessible to language processes. The result is memory-based mechanism that priorities information that has been in previous shared attention. We term this, 'cognitive common ground'. Cognitive common ground pre-dates the earliest development of Theory of Mind (at the end of the first year)[10] and can function independently of Theory of Mind abilities, relying solely at first on lower-level biases.[11]

In previous sections, we saw that shared, or joint, attention is also accompanied by these social ('ostensive') cues and that it can be shown that shared attention to some fact can serve as a basis for two people having common knowledge of that fact (Barwise, 1988/1989; Peacocke, 2005), where 'common knowledge' is as defined in Lewis (1969) and assumed in the work of Clark (Clark 1996) and elsewhere. As discussed in Peacocke (2005), Tomasello et al. (2005) and Carpenter & Liebal (2011), joint or shared attention relies on cognitive abilities for representing intentional action but may not call on the full range of theory of mind abilities. Thus, shared attention is not only a basis for common knowledge, it is premised on more minimal cognitive abilities than common knowledge.

The proposal then is as follows: there are cognitive mechanisms, that may not amount to more than processing biases, that are conditioned by certain social cues (eye-contact, contingent behaviour, self-directed utterances). These mechanisms are liable to produce a kind of proto-common ground independently of any cognitive abilities that are conceptually necessary for full-blown common knowledge. These biases are present in infants from a very early age and are conjectured to be innate. The situational cues are present in shared, or joint, attention and also early communicative or referential processes in infants. I conjectured

---

[10] In this thesis, it is assumed that infants from just after the end of their first year are able to predict other agents' actions based on a false belief (see Baillargeon & Scott 2010, Song & Baillargeon 2008; Southgate et al. 2010) and thus possess functional Theory of Mind abilities. We leave aside the question of the extent to which pre-schoolers are able to integrate Theory of Mind reasoning into everyday cognitive processes.

[11] The proposals here differ from two-level approaches to the functions served by ToM, as found in Apperly & Butterfill (2009). I do not assume here that there are proto-mental state concepts that track some but not all states classified fully fledged mental state concepts (although I do not rule it out). According to our view, concepts of shared attention track shared attention (an intentional activity) but in infants, those concepts are not associated with a conception of intention that allows for the full set of ToM inferences that link intentions with beliefs and desires.

that the apparent requirement that informative intentions be ‘open’ or ‘overt’ arises not because there is common knowledge of the informative intention but because there is shared attention to the informative act. It is of course possible to infer that a communicator’s informative intention is common knowledge on the basis of shared attention to the informative act (since shared attention to the informative act is a ‘reflexive common indicator’ of the informative intention), however this is not necessary and it is certainly not possible for infants before their first birthday, at least.

This proposal has consequences for how we think of common ground and what role it might play in cognition. Before spelling those consequences out, the next section considers how Clark and colleagues propose that common knowledge of discourse relevant facts is established in cognition.

5. Common knowledge in language processing

As mentioned, there is a long tradition in psycholinguistic research based on ideas put forward in by Herb Clark and colleagues (Clark & Marshall, 1978; Clark & Marshall 1981; Clark, 1996). According to this view, common ground is just the set of discourse-relevant pieces of information that are common knowledge and definite reference is rooted in what is common or mutual knowledge. From its earliest manifestation in Clark & Marshall (1978), there have been two important themes to this research. The first was discussed at length in previous sections and is the idea that common knowledge of a referent is necessary for interlocutors in order for successful co-ordination to be achieved, since otherwise interlocutors cannot be certain that they are thinking about the same referent. The second theme, also mentioned briefly above, has to do with how common knowledge is established and maintained. Clark & Marshall (1978, 1981) propose a number of heuristics that enable interlocutors to infer that a referent is common knowledge to a reasonable inductive standard. Clark & Marshall also note that referents established in communication being common knowledge require a different kind of memory representation to that standardly assumed in computational circles at the time. Standard assumptions held that an object in memory is represented along with a list of properties, facts about the object gleaned from experience. Clark & Marshall (1978) note that a referent's being common knowledge requires a kind of diary representation for objects in memory, encoding not only facts about that object but where given pieces of information were sourced; particularly through communicative or social interaction.

In previous sections we considered at length whether common knowledge of definite referents is in fact a necessary standard for interlocutors to be sure that they have coordinated on the same object. The answer to that is that it is not. While it is often necessary to think about our interlocutors' beliefs and expectations to establish a referent, common knowledge of a referent seems to be more a *result* of referential processes, rather than a pre-condition.[12]

In the last section, it was argued instead that the presence of common knowledge in social interactions that involve co-ordination is due to a prior more primitive state of *shared attention*, supported in turn by lower-level cognitive biases triggered by social cues such as eye-contact, contingent behaviour and self-directed speech. The standard references on common ground in psycholinguistics are Clark &Marshall (1981) and Clark & Marshall (1978). They both propose that common knowledge of a referent is established through various heuristics. These heuristics are inspired by Lewis (1969) and determine what Lewis terms a 'reflexive common indicator' for common knowledge. They fall into three broad categories: co-presence, previous discourse and background or general

[12] Note that, as discussed above, a cognitively realistic analysis of common knowledge should be understood in terms of potential mental states, rather than actual beliefs etc. Thus even if common knowledge, in this sense, is a product of referential processes, it does not follow that interlocutors make inferences about each others' higher order expectations once reference is already established.

knowledge. In this section we will consider all three heuristics and show that in fact common knowledge of discourse-relevant information may be merely a bi-product of the maintenance of a common ground by lower-level mechanisms for social interactions and by shared attention.

Let us first turn to co-presence heuristics. Clark & Marshall (1981) ask us to consider a scenario where two people face each other at a table with a candle standing in between. The idea is that the co-presence of the other agent together with salient object makes that object a candidate for common knowledge with that agent. However, Clark & Marshall are careful to stipulate that a co-presence heuristic would rely on shared attention to the salient object (for one of the two people may be lost in thought and not notice the candle in front of them). Thus co-presence heuristics really rely on there being shared attention to the object or event in the utterance situation. Thus the cues in question are really cues to shared attention to an object or event and agents need only be inferring shared attention of that object or event.[13] Once it is inferred that the candle is in shared attention, common knowledge would follow.

The second kind of heuristic relates to previous discourse. Clark & Marshall (1981) discuss the example of a speaker uttering a sentence containing an indefinite description, ‘a candle’. That utterance is able to introduce a new referent (the candle the speaker has in mind) into the discourse. Clark & Marshall’s proposed discourse-based heuristics simply exploit this kind of process to establish common knowledge of the referents so introduced. Again, from a more minimal perspective, we could say that utterances involve shared attention to the informative intention. Shared attention to the informative intention could subsequently lead to shared attention to the content of that intention. Thus, one could argue that the heuristic takes the interlocutors from the informative intention that is in

[13] To reiterate a point made above, according to most authors, shared attention involves abilities to represent intentional states, not merely to recognise behavioural cues to those states (these being the cues to shared attention). See Carpenter & Leibal (2011), Peacocke (2005).

shared attention to shared attention to the content of that intention and hence to the candle introduced by the indefinite description.[14]

The third heuristic relies on shared background or culturally available knowledge. Clark & Marshall (1981) use the example of a candle being introduced in discourse followed by the speaker producing an utterance containing the definite phrase, ‘the wick’. The idea is that speaker and hearer have shared background knowledge that candles have wicks. Heuristics based on such background knowledge allow that objects currently active in discourse can be linked to other referents via assumptions about the world that can be taken to be common knowledge. The problem with this heuristic is that it very often redundant, as where accommodation takes place in discourse. This is illustrated in some of the cheese examples discussed above and it can be illustrated with examples of bridging reference, due originally to Clark. For instance the sentence, ‘Bill checked the picnic supplies and discovered that the beer was warm’ can be used where it is only common knowledge that Bill has picnic supplies (not their contents). The point about this example is that it need not be common knowledge that picnic supplies can contain beer, before the utterance is made. Arguably, the process by which the presence of the beer in the picnic becomes common knowledge involves inferences along the lines described in the cheese examples above: the hearer infers that the speaker expects the hearer to make a connection between picnics and beer and see that the utterance is relevant and informative if this assumption is made. While such inferences rely on making assumptions about the speaker’s expectations, beliefs etc., and hence Theory of Mind, the goal is not to infer what might be common knowledge but what the speaker might reasonably have intended. That it becomes common knowledge that picnics contain beer, or that Bill’s picnics typically contain beer, is an effect of the utterance, often intended by the speaker.

[14] Note that the informative intention that the speaker introduce a certain candle into discourse is a mental state. Joint attention to that state is different from joint attention to the candle that is introduced.

To sum up, we have considered very influential proposals about how common knowledge of discourse-relevant facts is cognitively realised using heuristic processes. I have argued that when it comes to co-presence and prior discourse, what is established via heuristic processes is something closer to shared (or joint) attention to discourse-relevant facts. The third kind of heuristic links current referents to other referents via common background assumptions. Here I argued that this heuristic is largely ineffective since decisions about a referent in these cases rely more on inferences about the speaker's intentions, assuming they are being co-operative, relevant and so forth. Background assumptions typically emerge as shared only as a result of the process of reference assignment.

6. Proposal

Let us take a step back and consider our ideas of common ground and its role in language processing. It is often taken for granted in the psycholinguistics literature and in the formal semantics literature that common ground is common knowledge of discourse relevant facts (Clark, 1992; Stalnaker, 2002). Influential work on common knowledge (Barwise, 1988/1989; Peacocke, 2005) has argued that in many key areas of human social interaction, including communication, common knowledge emerges as a result of a more basic relation of shared, or joint, attention.[15] One can accept that it is possible to analyse common knowledge so that it is a cognitively realistic notion. This is done for instance in Lewis (1969) and Sperber & Wilson (1986). However, I have argued, contra Clark and colleagues, that common knowledge plays no effective role in the cognitive processes associated with coordinating on a referent for a definite description, although Theory of Mind abilities often do feature in such processes.

[15] These authors are careful to not claim that common knowledge can only arise through shared attention.

We then considered three kinds of heuristic that Clark & Marshall propose enable interlocutors to establish common knowledge of discourse-relevant facts. As an aside it should be noted that it is necessary to suppose such heuristics exist since it is not cognitively plausible to establish common knowledge of some fact by checking each of an infinite series of higher-order expectations, i.e. via Theory of Mind reasoning. It was argued that two key heuristics (based on co-presence and previous discourse) arguably establish shared attention to such facts, as a step on the way to common knowledge of such facts. The third heuristic (based on background knowledge) is often ineffective, relying on pragmatic inference about the speaker's informative intention based on ideas of co-operation and relevance to establish the background information as shared as a bi-product of referential processes.

At this point, it seems reasonable to dissociate the idea of common ground from common knowledge. In continuity with the development of cognitive processes in infancy, we should view common ground as that information which is, or has been, the target of shared attention. Cognitive mechanisms for common ground are sensitive to cues to shared behaviour, such as eye contact, contingent behaviour and self-directed speech and are likely to include simple biases that devote greater attention to information made manifest when such cues are present. Rather than being a necessary constraint on referential processing, common ground among interlocutors emerges as a result of such processes.

An observation to make here, that will be relevant below, is that cognitive mechanisms for common ground do not necessarily call on Theory of Mind inferences needed for inferences about higher-order expectations. In fact, Theory of Mind abilities are involved in establishing common ground only in as far as they are involved in the referential processes that feed into determining the pool of shared information in common ground, as in the cases of accommodation (the picnic example and some of the cheese examples above). Indeed, this is a point on which we can agree with Clark and colleagues. Even Clark & Marshall's heuristics for establishing common knowledge place virtually no burden on systems that might be involved in ToM inference. Thus the role of Theory of Mind abilities, the abilities for making inferences about higher order beliefs or expectations, about what one's interlocutor knows and does not know, or has a false belief about do not play a significant role in maintaining common ground, whether common ground is conceived of in terms of common knowledge or in terms of more basic shared attention and lower-level mechanisms. We shall return to this point below.

In the next chapter, we turn to consider recent work on co-ordination in conversations that takes a more mechanistic approach to how it is achieved. These proposals generally invoke domain-general mechanisms for memory and priming and may seem to be set against the spirit of social-cognitive approaches to conversation, such as that being defended in this thesis, and also against the spirit of Clark's alternative. The chapter will explore where the differences may lie.

Chapter 2: Memory-based mechanisms for common ground.

1. Introduction

In the last chapter, we focused on widely discussed ideas about how agents co-ordinate on referents in conversation. These ideas stem from the philosophical and linguistics literature and make a link between a notion of shared information, 'common ground', and common knowledge. We saw that 'common knowledge' as set out in Lewis (1969) is one particular description of how a co-ordination process could satisfy our intuitions that such processes are 'open'. We argued that there are more minimal accounts of openness in communication in terms of shared, or joint, attention (Barwise, 1988/1989; Peacocke, 2005) and that these more minimal accounts have a better fit with developmental data, which suggest that low-level mechanisms mediate infants' inferences about novelty and familiarity. The sketch provided in Chapter 1 introduced the idea that processing biases yielded relatively more attention to the contents of shared events than unshared events. My proposal in that chapter assumed the existence of memory mechanisms that use the identity of an agent as a cue to memories of past shared experience. In this chapter, we will review research that explores how memory mechanisms might function to promote the salience of previously shared experiences with an interlocutor. This research has largely been presented as evidence for the thesis that simple domain-general memory mechanisms are responsible for a kind of de facto common ground. Thus, memory-based models for common ground are similar to the one being proposed in this thesis. One important difference however lies in views of the domain-specifity of the mechanisms. According to the proposal I am defending here, memory based mechanisms are constrained by social-cognitive factors, to do with shared attention, and low-level social cues at the memory coding phase. This chapter will present two experiments that challenge the domain generality of memory-based mechanisms for common ground and provide evidence that these mechanisms are constrained by the relevant social-cognitive factors.

2. Shared information in memory and processing

In Clark & Marshall's proposals, the triple co-presence (the speaker, the addressee, and the referent) is saved in the memory system called referential diary. The diary is "our own personal experience, supplemented by cultural histories and atlases for cultural co-presence and by various reference texts for indirect co-presence" (Clark & Marshall, 1978). The interlocutors refer to the diary when they make and interpret utterances. In Clark & Marshall's theory, felicitous conversation based on the common ground is mediated by memory mechanisms that code separately the source of information, as well as the information itself. During a conversation, interlocutors trace their episodic memory to coordinate their language use.

In recent times, various memory-based models of co-ordination in dialogue have been proposed that may seem to be strong competitor to Clark's theory. Memory-based models stress the role of an emergent property of ordinary memory processes in language use. In this section, we will focus on work by Horton, Gerrig and colleagues which proposes that assessments of common ground are mediated via low-level automatic memory mechanisms, with an occasional strategic communality assessment (Horton & Gerrig, 2005a, 2005b; Horton, 2007).[16] In particular, a wide range of information, including the identity of the interlocutors, is encoded in the memory representation of an utterance event. During a conversation the identity of the interlocutors primes the related information in the episodic memory. A speaker plans their utterance based on the

[16] We acknowledge that similar proposals have been made by Garrod & Pickering (see Garrod & Pickering, 2009, Pickering & Garrod, 2004). These are similar in that the see low-level mechanisms such as priming as playing a significant role in co-ordination in conversation.

activated memory related to their interlocutors, and the audience comprehends the utterance based on the memory related with their interlocutors as well (Horton, 2007).

The memory-based model claims that the accessibility of the memory plays an important role in language use. Through analyzing the CallHome English Corpus, Horton and Gerrig (2005a) explored two processes in conversation: commonality assessment, in which information is retrieved from memory concerning what information is shared with an addressee; and message formation, which decides how to use that information in conversation. A well-designed utterance and accurate comprehension are established on a basis of memory representation. To be specific, when contextual cues largely overlapped with original encoding context, the memory retrieval of the overlapped information will occur to facilitate language processing. Horton and Gerrig (2005) explored this memory association effect. Here, participants were required to map cards with two persons. In an Orthogonal card condition each person was mapped to separated card categories (eg, person A mapped to all bird cards and all dog cards; person B mapped to all fish cards and all frog cards), in contrast to an Overlapped card condition where two persons were mapped to all the card categories (eg, person A mapped to bird card 1 and card 2, dog card 1 and card 2, fish card 1 can card 2, frog card 1 and card 2; person B mapped to bird card 3 and 4, dog card 3 and 4, fish card 3 and 4, frog card 3 and 4). In other words, in the former condition participants only had to remember the category-partner association, but in the latter condition they had to remember the comparatively heavy loaded item-partner association. After a few familiarization trials, a test phase happened with old combination trials (eg, person A mapped to bird card 1 and 2) and new combination trials (eg, in the Orthogonal card condition person A mapped to fish card 1 and 2, and in the Overlapped card condition person A mapped to bird card 3 and 4) (for details, see table 1). The results were consistent with the supposition that participants in the Overlapping condition would exert more effort, in that they showed a trend for a longer time to initiate their descriptions of new versus old cards. They also analyzed the reconceptualization that participants adjusted their previous expressions. For example, in this experiment a bird was described as “a perched bluejay” in one round, and “a bluejay that’s perched without the legs” in a following round. The latter expression was reconceptualized by adding “without the legs” to the former one. It was found that in the Orthogonal condition directors generated more reconceptualizations for new versus old cards than in the Overlapping condition, and it may be due to fewer memory burdens in the former condition (Horton & Gerrig, 2005a). The participants were supposed to be associated with each item equally if there was not any character-card association. However, the difference between the Orthogonal and Overlapping condition indicated that the different accessibilities of memory representation greatly influenced people's speech. Although this study is a more memory task than a dialogue task, the results revealed a speaker-referent associative effect, and it is possible that the effect is widely used in the real-world dialogue.

Table 2. 1. Sample card distribution in Horton and Gerrig 2005 study.

Sample orthogonal and overlapping distributions of cards matched by participant pairs in each round

	Current Matcher	Orthogonal Card Distribution		Overlapping Card Distribution	
Rounds 1–4	Matcher A	Bird 1 2 3 4		Bird 1 2	Fish 1 2
		Dog 1 2 3 4		Dog 1 2	Frog 1 2
		plus eight filler cards		*plus eight filler cards*	
Rounds 5–8	Matcher B		Fish 1 2 3 4	Bird 3 4	Fish 3 4
			Frog 1 2 3 4	Dog 3 4	Dog 3 4
		plus eight filler cards		*plus eight filler cards*	
Round 9	Matcher A[a]	Bird 1 2 3 4	Fish **1 2 3 4**	Bird 1 2 **3 4**	Fish 1 2 **3 4**
		Dog 1 2 3 4	Frog **1 2 3 4**	Dog 1 2 **3 4**	Frog 1 2 **3 4**
Round 10	Matcher B[a]	Bird **1 2 3 4**	Fish 1 2 3 4	Bird **1 2** 3 4	Fish **1 2** 3 4
		Dog **1 2 3 4**	Frog 1 2 3 4	Dog **1 2** 3 4	Frog **1 2** 3 4

Note. Each number following the card categories refers to a different category exemplar. The numbers marked in bold in Rounds 9 and 10 represent the cards that were coded as “new” for this configuration of cards.
[a] For Rounds 9 and 10, the order of the Matchers was counterbalanced across triads.

(Horton & Gerrig, 2005a, p. 132)

In the memory-based model the accessibility of memory representation is highly influenced by partner-based cues. “The ease with which speakers could access partner-specific associations would greatly influence the extent to which evidence for audience design would emerge in their utterances (Horton, 2007).” Speakers design their utterance based on the available memory primed by their audience, and the audience comprehend the utterance also based on the assumption of partner-based audience design. The stronger the association between their interlocutors and information is, the easier the retrieval of this particular information is. (Horton & Gerrig, 2005a, 2005b). For example, supposing a scenario in which interlocutors A and B are discussing an event E in which character A and B participated. The event E will be easily accessible when A and B subsequently interact, because the event E is saved in an episodic memory with the character A and B together, and the identity of A and B activated the memory of the event E. But when A and another interlocutor C discuss the event E, the memory representation of the event E will be less accessible, because the conversational partner B provides stronger cues than C to retrieve the elements in this event. This character-based memory effect can also be found in different forms of language processing. In text processing, characters in the text also serve as memory cues to information associated with them (McKoon & Ratcliff, 1992; Greene, Gerrig, McKoon, and Ratcliff, 1994; Rawson, 2004). In broken precedent research in which a familiar precedent is substituted by a novel expression in a following conversation, participants are usually slower to respond when the same speaker uses a novel expression, compared with a different speaker (Metzing & Brennan, 2003; Kronmüller & Barr, 2007; Matthews, Lieven & Tomasello, 2010).

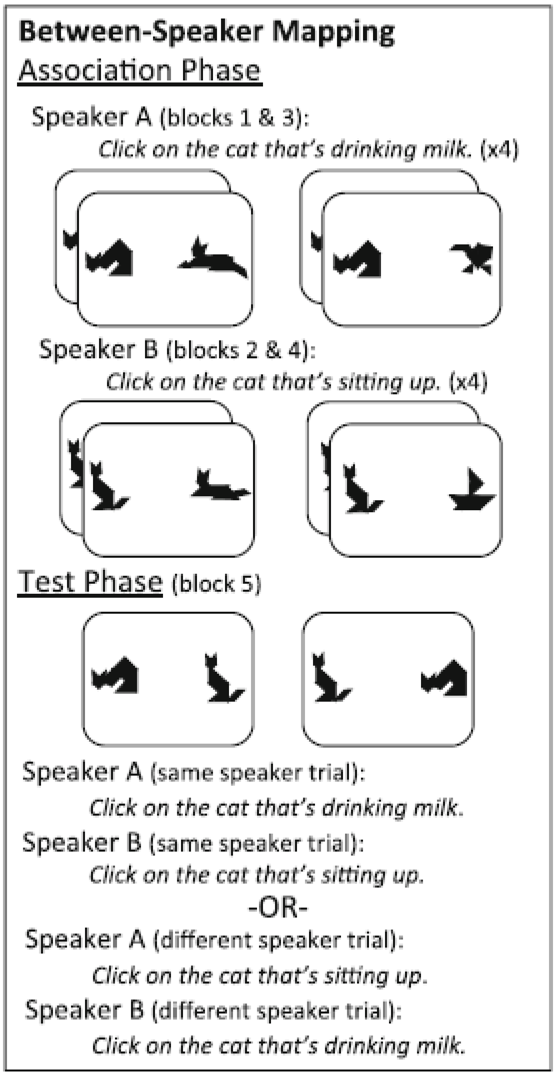


Figure 2. 1. A sample display of Horton and Slaten 2012's study.

This chart illustrates one of the conditions: between-speaker condition, where two speakers introduced two different pictures. The study examined whether participant linked the pictures to specific speaker in the association phase. (Horton & Slaten, 2012, p.115)

It has also been proposed that this cue-based system functions on a basis of domain-general memory processing. Regarding the most influential cue – the speaker-specific cue, memory retrieval comes from the compound voice and visual property of specific interlocutors. Horton and Slaten (2012) explored whether speaker-specific information facilitated participants' anticipation for a potential referent. In their task, in a first association phase two different pictures were associated with each speaker (e.g., speaker A with "cat that's drinking milk" vs. speaker B with "cat that's sitting up"). In a following test phase participants were presented the same sentence as the previous phase (e.g., "click the cat that's drinking milk"). They analyzed participants' fixations on the target "cat" before the disambiguous phrase "that's drinking milk" unfolded. In a same-speaker condition the following phrase referred to the old matched object as the first phase (e.g., A with "cat that's drinking milk). In a different speaker condition it referred to the new switched object (eg, A with "cat that's sitting up") (see figure 2.1). The results revealed a partner-facilitating effect in that participants largely anticipated the old matched picture when they heard the critical ambiguous noun. In the first experiment two speakers were in different genders, so it was possible that the shared knowledge (gender) was the cue for memory retrieval. It is noteworthy that in the following experiment with the same setup, when two speakers had the same gender, participants still demonstrated the same performance pattern as the first experiment. It indicated that it was the speaker-specific cues, such as the voice and facial information, that primed the encoded information. It indicates that the gender information did not play a determinative role as the specific property of the speaker cue did, even though the gender information may assist processing the specific speaker cues.

Memory-based theory suggests that low-level ordinary memory processes mediate common ground, rather than something more like explicit memory. The main evidence comes from a lack of correlation between an implicit partner-specific memory and an explicit memory recognition task (Horton & Gerrig, 2005b; Horton & Slaten, 2012; Horton, 2007). In Horton's 2007 first study, participants attempted to generate exemplar word (e.g., banjo) under directors' category hints (e.g., a musical instrument). Two directors assigned different items (e.g., director A assigned items "banjo"; director B assigned items "harp") in the same category. In a following picture naming task, item pictures were allocated by either the same director (e.g., director A assigned items "banjo"), or a different director (e.g., director A assigned items “harp”), or not mentioned (e.g., director A assigned a new item) (see Table 2.2). They found a shorter naming latency for items with the same partner compared to with the different partner, and revealed the partner-based memory influenced the language production. In a last partner identification task participants attempted to identify whether the item appeared in the first task and with which director it was associated. No significant correlation was found between the explicit partner-specific memory in the final task and partner-specific priming in the picture naming task, which indicated that common ground is applied by low-level memory mechanisms (Horton, 2007). Tasks such as these demonstrate that participants retrieved their memory primed by a specific partner, but their performance of specifying the specific partner in an explicit recognition task is entirely unrelated (Horton & Gerrig, 2005b; Horton & Slaten, 2012).

Table 2. 2. Sample task phases and items in Horton 2007 study.

	Partner context	
Task phase	*Partner A*	*Partner B*
Phase 1: Exemplar Generation	Cue: *'a musical instrument'* Fragment: **B _ _ J O** Cue: *'a bird'* Fragment: **O S _ _ I C _** Cue: *'an occupation'* Fragment: **D _ N T _ S _** *Etc.*	Cue: *'a musical instrument'* Fragment: **H _ R P** Cue: *'a bird'* Fragment: **P _ N G _ I _** Cue: *'an occupation'* Fragment: **_ U R _ E** *Etc.*
Phase 2: Picture Naming[a]	*Ostrich* (Same partner) *Harp* (Different partner) *Cactus* (Novel control) *Etc.*	*Penguin* (Same partner) *Banjo* (Different partner) *Skis* (Novel control) *Etc.*
Phase 3: Partner Identification[b]	**DENTIST** ('old'-Partner A) **EARRING** ('new') **HARP** ('old' – Partner B) *Etc.*	

[a]Item names listed in Phase 2 represent pictorial stimuli.
[b]Participants carried out Phase 3 in isolation, independent of either partner context.

(Horton, 2007, p.1120)

In summary, the kind of memory-based model proposed by Horton and colleagues sees a key process for recovering shared information as being based on the memory system. This could be seen to contrast with a model built out of Clark & Marshall's co-presence heuristic, whose diary-like representations may be seen to be explicit. Instead the memory-based model emphasizes low-level automatic memory associations. In this model, an automatic cue-priming process happens in a retrieval process, after a wide range of contextual information is resonant with an encoding process, and the whole process is implicit. Besides, what determined the activation of the memory representation is the contextual cue. When it comes to the cue, the most influential one is the identity of the conversational partner. A conversational partner may activate a wide range of information through a domain-general memory association. The extent to which information is employed is determined by the salience and relevance of speaker cues. Therefore, when speaker-specific information is strong enough to pass a threshold of activation, addressees will employ the activated memory representation to make an inference. Similarly, the speaker includes audience design in his speech based on the partner-cued memory representation. This fast implicit calculation contributes to the successful co-ordination in conversation.

In contrast to how low-level memory mechanisms might make available potentially shared information, the standard interpretation of Clark & Marshall's proposals suggests a more resource-intensive and explicit process. In Clark's co-operative heuristics, interlocutors have to establish a rich common ground regarding what the interlocutors commonly know, what they said, what they can see, and so on. All that information is saved into a rich and diary-like episodic memory. To retrieve the information, people have to retrieve information from their episodic memory. To maintain a felicitous conversation, people need to keep checking the memory reservoir. But considering the time pressure of the dialogue, it is hard to examine all the information in online language processing.

In contrast to the standard interpretation of Clark & Marshall's suggestions, we considered an alternative view of common ground in earlier sections which sees social cues and shared attention to previous utterance content as providing the basis of a record of previous shared information. The proposal relies on the idea that, relatively speaking, more attention is devoted to information in the 'shared situation' created by shared attention. This information includes who the other agent is and what is attended to. In terms of verbal communication, this translates as who the speaker is and the content of previous utterances.

Without the constraint of shared engagement, people can establish unlimited association between a token and contexts, which is unrealistic for the human being. Hence, the conversational partners can be considered as one of the most important contexts for establishing common ground, because they are easily distinguished from the other domain-general contexts due to their social property (Brown-Schmidt, 2015). The attentional relevance from the conversational partner determines which information is encoded and to which extent it is encoded. The encoded information is represented as a gradient way depending on the salience and relevance of shared information (Brown-Schmidt, 2012). The more salient and relevant the partner-associated information is, the stronger the association is formed.

In sum, the memory-based model proposes a low-level processing, which means that the processing is automatic and cognitive resource free. It does not propose that people necessarily process the rich and diary-like memory, as Clark and Marshall suggested. It also does not propose that people need make an inference of mental states. It indicates that perspective-taking occurs in an ordinary memory processing during which the related information is activated by the contextual cues. In this way the extent to which the shared information is used relies on the weight of the cue-based memory.

What evidence supports the low-level process? In Horton and Gerrig’s (2005a) study and Horton and Slaten’s (2012) study, participants all demonstrated a speaker-specific memory facilitated their referential anticipation in the tasks. However a following explicit memory task did not show any correlation between their performance on the referential task and the memory task. Participants might activate speaker-related referents without explicitly remembering the speaker-referent association. These results indicate that memory encoding and retrieval all occurred in an implicit way.

3. Speaker-specific cues in memory retrieval

In memory-based models speaker-specific cues play a vital role. The memory-based model suggests associations made in memory between interlocutors and specific events become re-activated in subsequent interactions. Although it is proposed that this is a result of domain-general processing, the speaker is the most influential cue in the memory. This speaker-specific cue is also appropriate to explain the complexity of the perspective-taking, based on the salience of the cue. However, it brings some challenge on the memory-based model.

First, the evidence did not differentiate the effect of common ground and the speaker-specific cue. For example, in Horton and Slaten 2012’s study, a speaker A was associated with the description "the cat that's drinking milk” in an association memory stage. In the following test stage participants demonstrated a speaker-specific performance, that is, when they were having a conversation with the speaker A, the memory representation of "cat that's drinking milk” was automatically retrieved. They largely anticipated the old match – speaker A associated with "cat that's drinking milk”. The evidence revealed that the speaker is a cue for related memory retrieval. However, it is also possible that the referent "cat that's drinking milk" became in the common ground between participants and the speaker A in the test stage, in light of the models proposing the common ground as a specific entity. It is not clear to distinguish whether the common ground or the speaker cue results in the anticipation of the target referent. Therefore, how information is encoded in memory and retrieved in the processing of perspective taking need to be further explored.

Second, Horton's memory-based model is mostly based on the non-conversational memory task, although participants attempt to formulate utterance at the word level. As they suggested that the memory representation is an ordinary property encoded and retrieved in a same manner as other cognitive processes, the experiments they conducted are more like a memory task than a language task. In this way, the evidence supporting the view that memory representation in the language use is implicit and automatic could have better foundations. Putting the model in conversation, it is sometimes hard to distinguish the speaker-specific association and the referent-expression memory. For example, Kronmüller and Barr (2007) showed that previous expressions (conversational precedents) prevented people to map new expressions to an old familiar referent. Listeners used the speaker information to inhibit the preempted expressions -- the speaker effect. But this speaker effect happened later and thus it is a distinct process to the precedent recovery. So whether the speaker-specific effect is generally applicable to language processing need to be further explored.

Regarding the first point, a way to differentiate the speaker-specific effect in a memory-based model and the constraint to Clark's common ground is to manipulate the speaker-specific effect on a basis of consistent common ground. As will be reviewed in the next chapter, prominent models used to explain behaviour in perspective-taking tasks predict that participants should demonstrate a binary performance depending whether the common ground information is accessed in referential decision processes. For example, in the perspective-adjustment model proposed by Keysar and colleagues (Keysar et al., 2000) people take all information available to them into account first and adjust later. In other models common ground knowledge can be processed at first. It will be hard to explain the evidence that under the same background people demonstrate different performance, depending on memory associations cued to the speaker. That is to say, when interlocutors shared the same information, they should be consistent in perspective taking depending on the assessment of the shared knowledge. But the memory-based model predicts that if the participants associate privately viewable objects with the speaker, this could affect on-line decision processes. In the next sections I will present our first set of experiments on this hypothesis. Additionally, the studies to be reported below will present a visual-world conversational task. This offers an advance in terms of evidence for memory-based models since most of the evidence from Horton was based on non-linguistic memory tasks or other offline tasks.

4. Experiment 1

This section will present the first of two experiments exploring perspective taking when the common ground keeps consistent across two interlocutors. In the present study, a visual-world experiment is based on the design in Keysar et al (2000) and many subsequent studies. Participants are asked to move objects around a 3*3 grid by an instructing speaker while being eye-tracked. Three of the nine grid positions are boxed off on one side so that objects in those positions are only visible to the participant. The other six positions are in common view. See Figure 2.2.

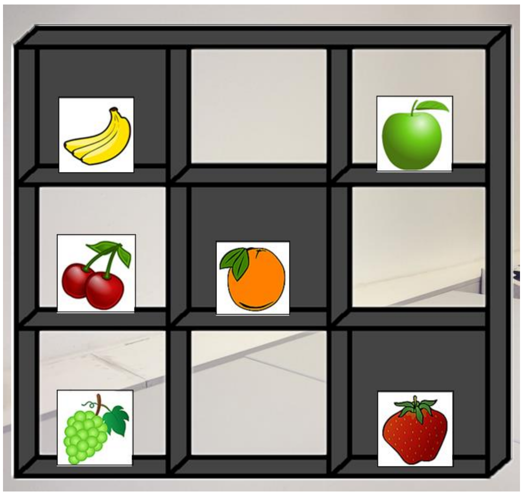


Figure 2. 2. The example display for the main game grids in Experiment 1.

Participants can see any objects in all the nine positions. The instructor sees the grid from the opposite side and the three objects in the private grey grids are not visible (here, the banana, the orange, and the strawberry).

Three of the common positions contain commonly viewable objects and the other three are empty. As in previous studies, critical trials involve an instruction like, ‘Move the apple to the bottom middle’ where the target object (in this case, the apple) is in common ground. In one condition, there is another apple (a competitor) among the three objects viewable only to the participant (in privileged ground). In another condition, none of the three privileged-ground objects is of the same type as the target (non-competitor distractors). As is common in these studies (see Ferguson & Breheny, 2011), the dependent variable is a log-transformed measure due to the fact that analyzing eye-gaze data directly on a proportion scale can lead to a distortion of effects (Barr, 2008). Thus a target advantage score which is the natural log of the proportion of fixations to the target divided by the proportion of fixations to the competitor/non-competitor distractor. I.e. DV = ln(P(Target)/P(Comp). In previous studies, a smaller target advantage score has been found in early time windows for the competitor condition compared to the distractor condition - suggesting that participants are not entirely able to ignore an object that fits the referential description even though it is not in common ground (it is in privileged ground). Let us call a smaller target advantage score for the competitor condition, competitor interference. According to memory-based accounts mentioned above (Horton, 2007; Horton & Slaten, 2012) we should expect to find a greater competitor interference to the extent that participants associate privileged positions with the speaker.

To test this hypothesis, we use two ways of referring to objects in the grid. In previous studies, reference is normally made using type-based descriptions. I.e. descriptions make reference to the type of object ('the apple', 'the dog', 'the triangle'). Another way to refer to objects in the grid is by a location-based description. For example, 'the object in the top right position'. Using location-based descriptions, unlike type-based descriptions, a speaker can give an instruction about an object that they cannot see. For example, the speaker could refer to the banana in Figure 2.2 as, 'the object in the top left position' while it is common ground that the speaker does not know what that object is. In our study, 24 filler instructions use location-based instructions. However, in one condition, all instructions refer only to objects in common ground - this includes location-based filler instructions. In another condition, two thirds of location-based fillers make reference to privileged ground objects. We call the common-ground only condition the *homogeneous condition*. We call the condition where the speaker gives instructions about both common ground and privileged ground objects the *heterogeneous condition*. In both conditions, critical items are the same and they all involve type-based descriptions of objects in common ground, as in previous studies.

A prediction of the memory based account is that we will see more competitor interference in the heterogeneous condition than in the homogeneous condition. This is because in the heterogeneous condition, the speaker refers to privileged positions and so a greater association will build up in memory between the speaker and privileged positions than in the homogeneous condition where the speaker does not refer to the privileged ground positions. This should cue the participant to attend to the privileged areas while being given instructions.

According to a domain-general, memory-based account, we can operationalize this design in a number of ways, as long as the contingencies that are available in the context of the experiment provide the same strength of cue for the participant. To test this hypothesis, we introduce a third variable which manipulates the type of contingency. In one condition, we use two speakers (one male, one female). One speaker gives instructions for the homogeneous condition and one gives instructions for the heterogeneous condition. In a second condition, only one speaker gives instructions but each instruction is accompanied by one of two sets of physical cues (a colour and a sound). For the homogeneous condition instructions, the participant sees and hears one set of colour and sound, for the heterogeneous condition instructions, the participants sees a second set of colour and sound. The prediction is that, to the extent that the contingencies in the heterogeneous conditions would provide an equally reliable cue for memory retrieval, there should be no effect of cue type.

As will be discussed at greater length in the next chapter, widely discussed models of perspective taking (such as perspective-adjustment models and constraint-based models) do not make a prediction about the effect of manipulating association (homo- vs. heterogeneous instructions). However, without any further assumptions about low-level memory factors, these accounts predict no difference between association conditions.

4.1. Method

4.1.1. Participants

43 participants (8 male and 12 female in the human cue condition; 9 male and 14 female in the physical cue condition) were recruited from UCL's participant pool. Two UCL students were recruited to play the speaker role. These students were blind to purpose of the study. All were paid or given course credit for their participation.

4.1.2. Stimuli and design

Participants were given instructions to move objects around a 3*3 grid as in Figure 2.2 above, where three objects were in common ground between the instructing speaker and hearer and three objects were placed in grid positions so that they were only visible to the hearer (privileged ground). Except for the training phase where participants were given a turn at being director, all instructions were given by a confederate speaker. These instructions were pre-recorded in a single session with each of the two confederates who worked on the study. Thus each participant in a given set of conditions heard the same pre-recorded instructions.

For each trial, a new set of objects appeared on the display. Each such set contained objects belonging to one of three categories: fruit, animals or geometric shapes. There were ten different pictures for each category and some were paired in the same subtypes. For example, among the fruits, there were two apples, two bananas, and two oranges. Six pictures were randomly chosen for each trial of the experiment, except for the competitor condition where two pictures from the same subtypes were presented. For instance, competitor condition item might contain a picture of a banana, cherries, grapes, an orange, and two different apples, while for the matched non-competitor condition one of the twinned objects was replaced by a new object, for example, a strawberry (Fig. 2.3).

A trial consisted of one instruction where the participant was asked to move an object to a location. For example, 'Move the apple to the bottom middle'. There were 24 critical trials and 24 filler items in each association condition. In 12 critical trials (competitor condition), the privileged objects included one that was the same type as (although not visually identical to) the target in the instruction (an apple). In the other 12 critical trials (non-competitor distractor condition), no privileged object was of the same type as the target object. Figure 2.3 below shows initial set ups for a pair of trials in competitor and non-competitor conditions.

Figure 2. 3. The initial display for the competitor condition (left) and non-competitor condition (right) in Experiment 1.

In this example, the target green apple had competitor with the same subtype, e.g. a red apple, (competitor condition) or an irrelevant non-competitor, e.g. a strawberry, (non-competitor) in participants' privileged grid.

The 24 filler items involved location-based instructions, 'Move the one in the bottom left to the middle right'. In a homogeneous association condition, all 48 instructions were targeted at objects in common ground. In a heterogeneous condition, 18 of the 24 location-based filler instructions were targeted at privileged ground objects.

Association was operationalized in two different ways in a between groups cue-type manipulation. In a social-cue condition, two different speakers (one male, one female) gave instructions in the same session. One speaker gave instructions that referred to common ground objects only (homogeneous condition) and one speaker gave instructions that referred to both common and privileged ground objects (heterogeneous condition). The gender of the speaker in each condition was counterbalanced across the experiment. In a physical-cue condition, one speaker gave all 96 instructions for the session. For half the instructions, the frame showing the speaker had a yellow border and individual instructions were preceded by a ringing sound, while half the instructions had the speaker's image in a green border and a distinctively different buzzing sound preceded them. One set of these physical cues accompanied homogeneous instructions while the other set accompanied the heterogeneous instructions. These were counterbalanced across the experiment.

To sum up, the experiment had a 2 (competitor) * 2 (association) * 2 (cue type) design. Altogether, each session with a participant involved 96 trials, 48 delivered in the homogeneous and 48 in the heterogeneous condition. The order of presentation of trials was randomised. Cue-type was a between subjects manipulation. So a participant either had instructions from two different people (social-cue condition) or one person but with two sets of colours and sounds (physical cue).

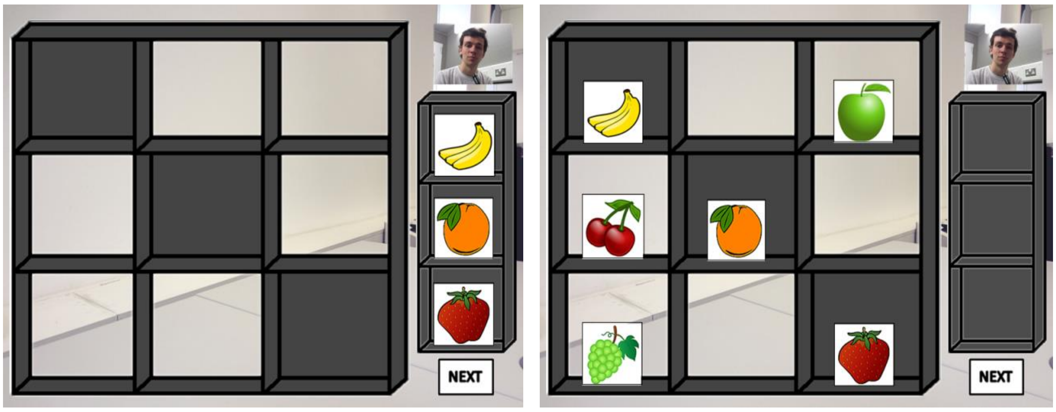
4.1.3. Procedure

In each session, a participant was seated in front of a 17” screen fitted with Tobii X60 eye-tracking equipment in a room with only the experimenter. They were told that they would be interacting with a person located in another part of the college via a video link. Prior to a training phase, participants were given some initial instructions. All instructions were given verbally by one of the confederate speakers via video appearing on the monitor in front of the participant. The speakers told the participants that they would play a game where they should follow the speaker’s instructions to move objects around a grid. They were then shown how the screen would be set up at the beginning of each trial. See Figure 2.4.a.

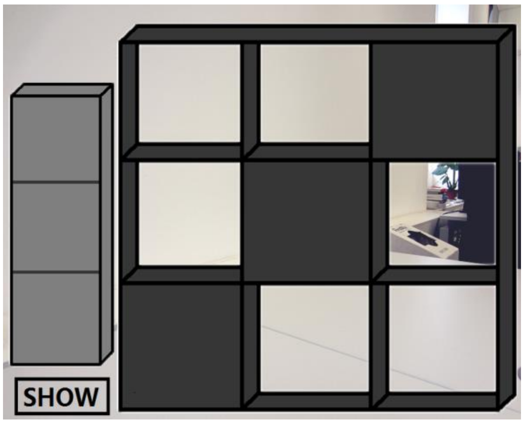
In figure 2.4.a, the 3*3 grid appears on the left of the screen. It is empty. On the right there is a shelf which contains three objects. In the training phase at the beginning of a session, the speaker tells the participant that these are their private objects that she/he does not know what they are or care what they are. The speaker instructs the participant that they should move these objects to the private positions on the 3*3 grid and then click on the image of the speaker in the top right hand corner of the screen when they are ready. When the participant has completed this task, they click on the image of the instructor and a bell sound is hear. The instructor then makes three objects appear in common view on the grid. The visual display now appears as in Figure 2.4.b. The instructor then gives one instruction to move an object to a new location. When this is completed by the participant, they click the ‘NEXT’ button on the bottom right of the screen and the whole procedure begins again for the next trial.

In the next part of the training phase, the participant experienced a few trials from the instructor’s perspective. They saw the empty grid with three blocked-off shelves, as in Figure 2.4c. Then they heard a bell sound and clickd “SHOW”. Then the common objects appear. One object is marked with a red star and one of the empty grid positions is marked with a blue star. Participants are instructed to tell the other person to move the marked object to the marked empty position. This part of the training procedure ensures that participants understand that the instructor does not see the common objects.

The participant then returns to the role of the listener and the experiment begins. Note that, on each new trial, the three private grid positions appeared in different, randomly assigned locations around the nine frame grid. After every two or three trials, participants had an opportunity to break prior to re-calibration of the eye-tracking equipment. Note also that the trial order was fixed across both groups. It was constructed using a randomised order except that three location-based fillers were placed among the first ten trials, prior to the occurrence of the first heterogeneous test trial.



a. b.



c.

Figure 2. 4. The initial procedure for each trial in Experiment 1.

Participants were firstly required to move three objects from their private small grid to the three private grey frames in the big grids (a), then three common objects were shown in the transparent frames (b.). A prior training phase involved the participant taking the instructor's role. They see the grid as in (c) initially prior to hearing a bell sound, at which point they click, 'SHOW'.

4.2. Analysis and Results

4.2.1. Data Processing

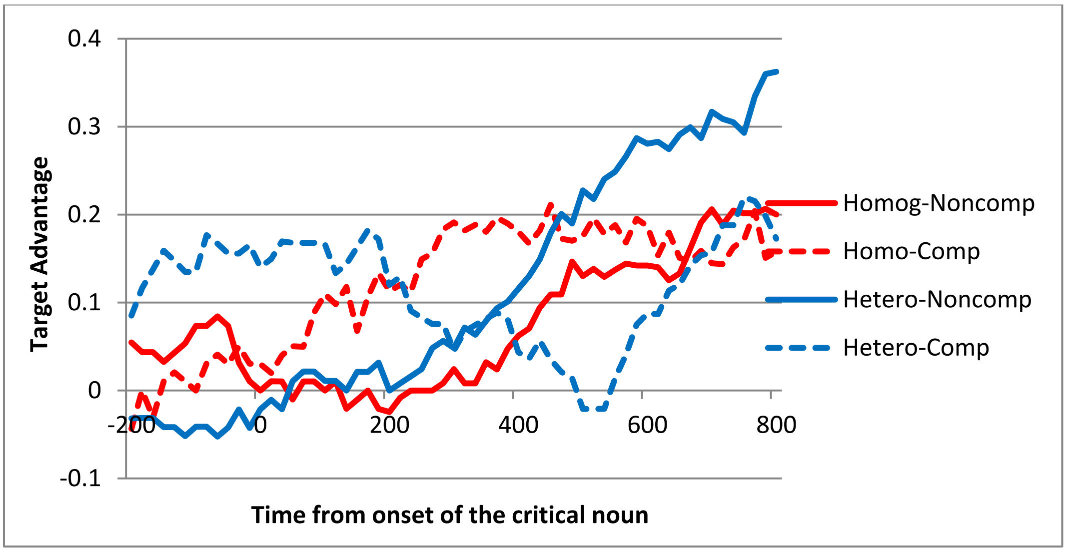
Eye-movements that were initiated during the auditory instruction were processed according to the critical word ('apple') onsets for the purpose of aggregating the location and duration of each sample from the eye tracker. For analysis, we removed any sample that was deemed 'invalid' due to blinks or head movements. The spatial coordinates of the eye movement samples (in pixels) were then mapped onto the appropriate object regions; if a fixation was located within the square surrounding an object, it was coded as belonging to that object, otherwise, it was coded as background. Target and competitor items were identified on a by-trial basis.

Probabilities of fixating the *target* or *competitor/non-competitor* objects as a function of time were analysed in terms of a target advantage score ln(Ptarget/Pcompetitor). Target advantage scores were analysed for three time regions. A preview region began 200ms before, and ended 200ms after, the onset of the critical word (‘apple’). The preview region extends to 200ms after the critical word onset in accordance with standard assumptions about the time to program and launch an eye movement (Hallett, 1986). Following the preview region, our analysis looks at two 300ms regions, 200-500ms and 500-800ms. Following Hanna et al., (2003) the overall 600ms region of interest was chosen to correspond to the average critical word length (between five hundred and six hundred milliseconds). After this period participants' eye movements were often directed toward the mouse curser. The regions were identified and synchronised for each participant on a trial-by-trial basis, relative to the onsets of the critical word.

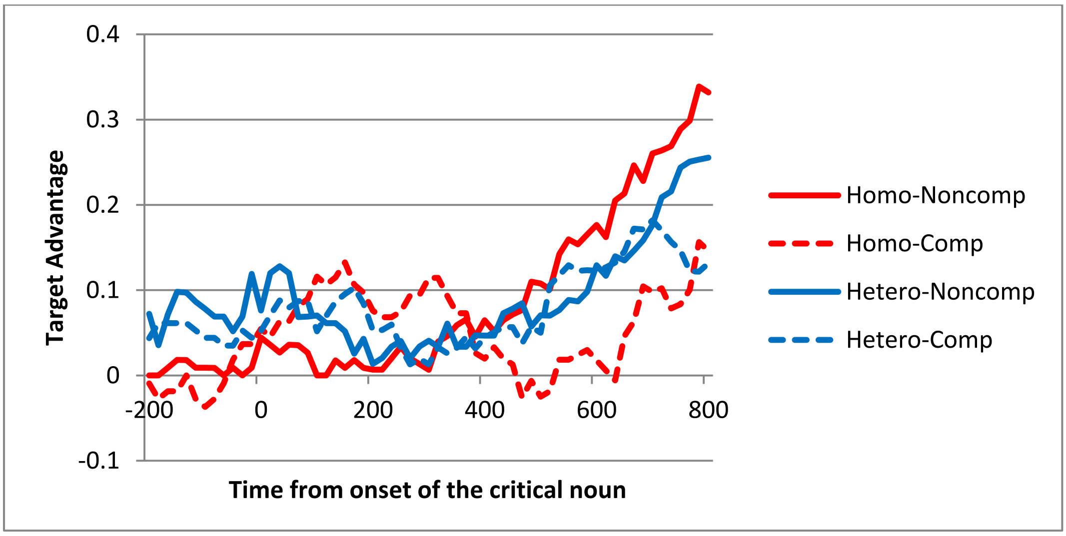
Because this experiment is designed to establish memory-based effects, we expect these to emerge over time. That is, any association the participant makes with a given cue and the private grid positions are likely to build up over time. Thus the ANOVA analysis includes an additional factor, Block. This is simply a division of the trials in each condition into those occurring in the first half and those occurring in the second half.

4.2.2. Results

For each participant (respectively item) and condition, I calculated the average target advantage score over the 17ms time slots per analysis region. Figure 2.5 shows a plot of the time course of target advantage for the two groups. Each graph of Figure 2.6 shows the average target advantage score for each of the competitor and non-competitor conditions for each of the three windows. And Figure 2.7 shows the proportion of fixations to the target and the competitor in eight conditions.

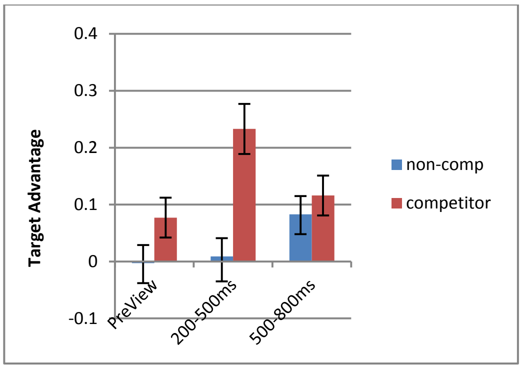


Human cue condition

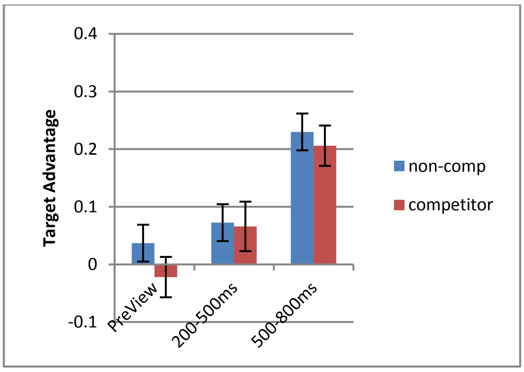


Physical cue condition

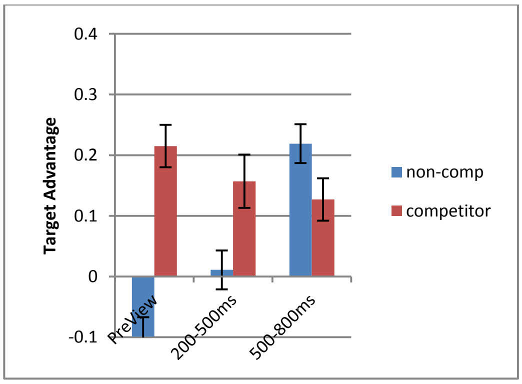
Figure 2. 5. The time course of the average target advantage score in Experiment 1 showing the TA score in the human cue condition (upper) and physical cue condition (lower) from the 200 ms before the onset of the critical word to 800 ms after the onset.



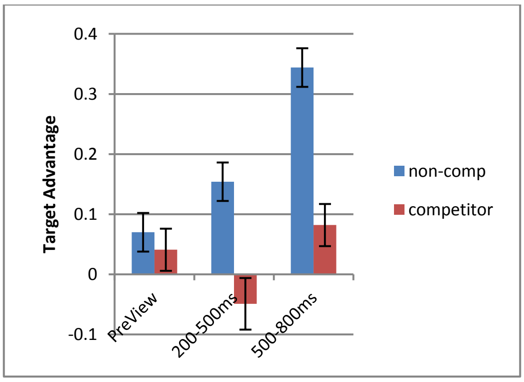
First block - Homogeneous



Second block – Homogeneous

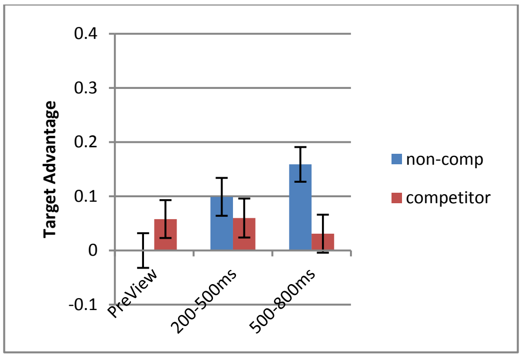


First block - Heterogeneous

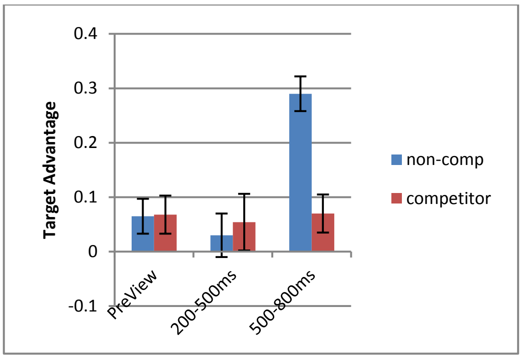


Second block - Heterogeneous

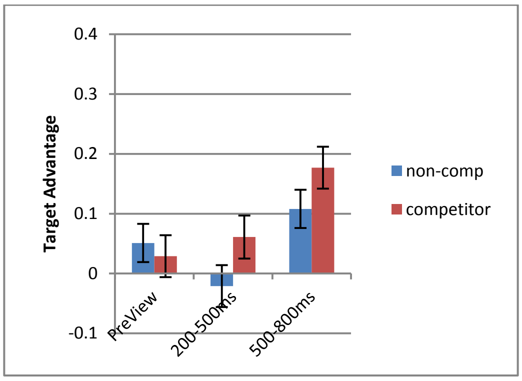
a. Human condition



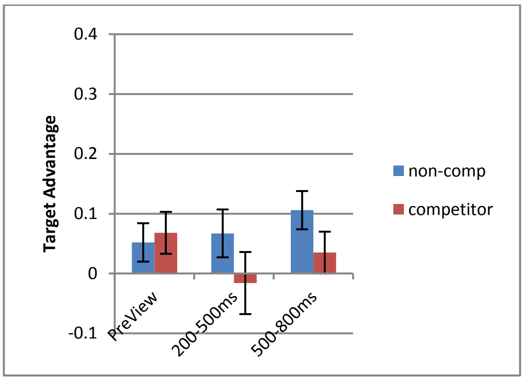
First block - Homogeneous



Second block – Homogeneous



First block - Heterogeneous

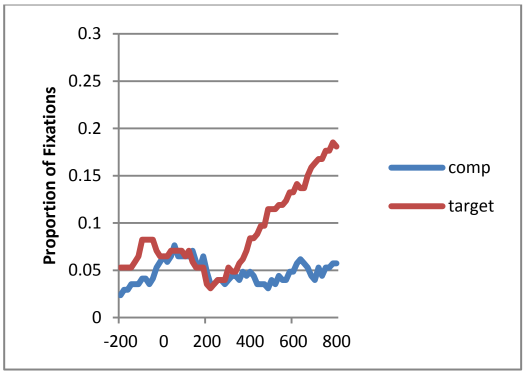


Second block - Heterogeneous

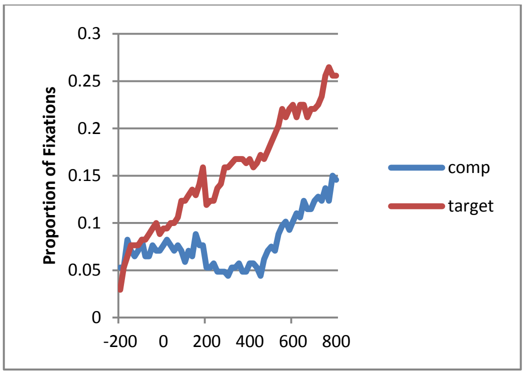
b. Physical condition

Figure 2. 6. The average target advantage score for the three time windows in Experiment 1 by block.

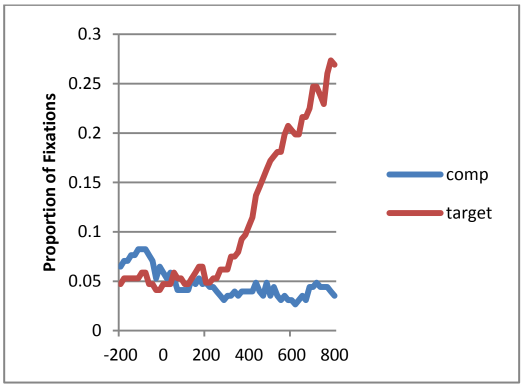
The charts show the competitor and non-competitor conditions in the human cue (2.6a) and the physical cue (2.6b) conditions. In each case, the averages are shown for the first block (left) and second block (right) for the homogeneous (upper) and heterogeneous (lower) conditions. Error bars indicate standard error of by-participant means.



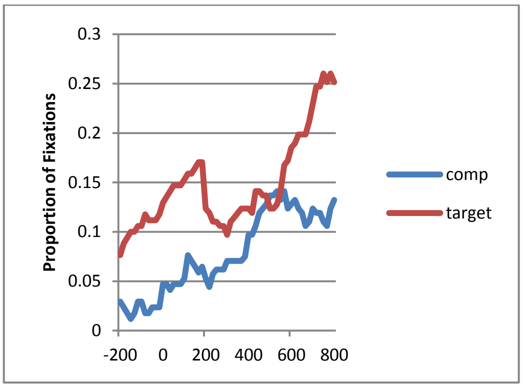
Homogeneous – Non-competitor



Homogeneous – Competitor

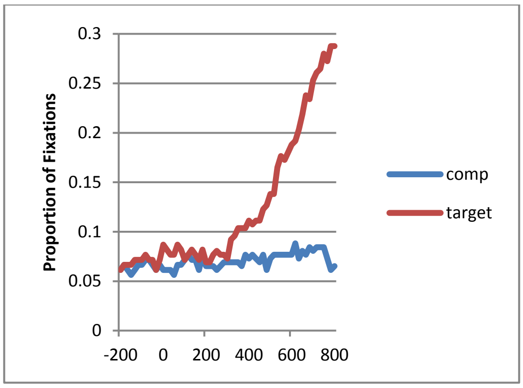


Heterogeneous – Non-competitor

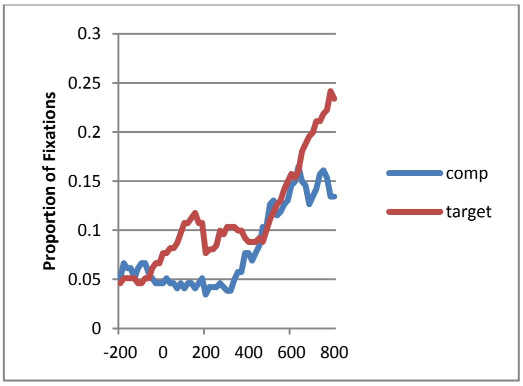


Heterogeneous– Competitor

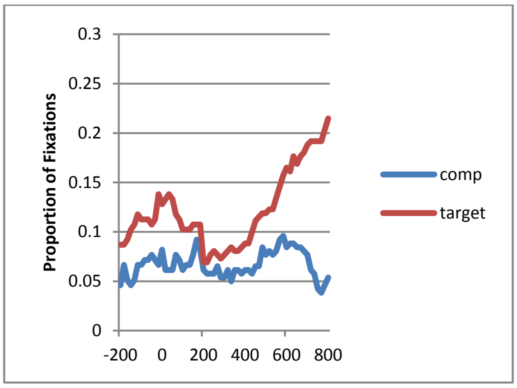
a. Human condition



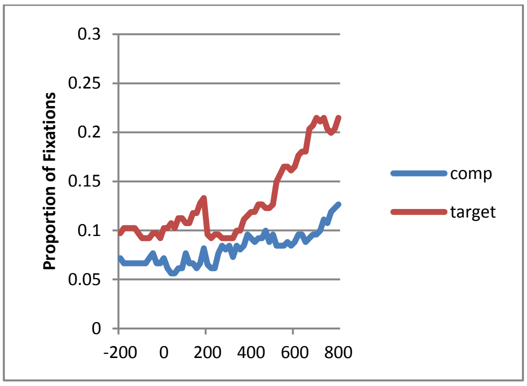
Homogeneous – Non-competitor



Homogeneous – Competitor



Heterogeneous – Non-competitor



Heterogeneous– Competitor

b. Physical condition

Figure 2. 7. The proportions of looks to the target and the competitor in Experiment 1.

The charts show the proportions of looks in the homogeneous-noncompetitor, the homogenous-competitor, the heterogeneous-noncomeptitor, the heterogeneous-competitor conditions in either the human (first four) or the physical (last four) conditions.

The first thing to notice about these graphs is in Figure 2.5, which shows the average Target Advantage scores plotted over time. In the physical cue condition, the overall pattern seems to be what one would expect. Toward the end of the time window (i.e. around 200ms after the offset the target word), target bias seems greater in trials where there is no competitor in private view, for both Homogeneous and Heterogeneous conditions. However, in the human cue condition, only in the Heterogeneous condition is the pattern as one would expect, with non-competitor trials attracting more bias than competitor. By contrast, in the Homogeneous condition, bias in the Non-competitor condition looks lower and roughly equal to the Competitor. To get some understanding of this unexpected result, we turn to figure 2.6, showing the results split across the two halves of the experiment. It is noticeable that in the first half of the human condition, target bias seems greater in the conditions where there is a private-view competitor, compared to where there is not, although this bias peaks before the final window. Correspondingly, the Target Advantage when there are Non-competitor trials seems very low, even though there is no competitor in the private view. It is easy to see both of these results as due to participants being strategic, particularly in the early part of a session. Since 24 out of 38 dual-objects trials were experimental trials, it is understandable that an expectation arose in participants when they saw two objects of the same kind, to expect the common view one of those to be the target. That this pattern emerged more in the first half of the session, suggests that it was a more conscious strategy. It seems participants 'figured out' the pattern after a few trials. Likewise, when a display did not have pair of equivalent objects, participants seem to have adopted an early strategy to attend to the private objects in the expectation that the target would be there. This strategy would have been reinforced by the fact that three private-target fillers were included in the first ten trials. It happened that these did not have dual objects.

It is interesting that the strategic behaviour in the first half of the session did not occur in the Physical cue group nearly so much as in the Human cue group (even though item order was the same for both groups).

If this observation is borne out statistically, it would suggest already an effect of cue type. We return to the point below.

The effects of these strategies seem far less prominent in the second half of the session where the time course of gaze formation seems to follow along expected lines. Nevertheless, the strategies of the first half of the sessions would be expected to result in relatively lower overall Target Advantage in Non-Competitor trials where they is most in evidence. This was in the Human condition, particularly the Human Homogeneous condition.

The results of a 2(competitor vs. non-competitor) * 2(homogeneous vs. heterogeneous) * 2(human vs. physical cue) * 2(first block vs. second block) mixed ANOVA are as follows.

Preview Window

In the preview window, there were no main effects, but a significant interaction between competitor and blocks, F1(1,82)=5.248, p<0.05, η_p^2=0.060, F2(1,20)=4.413, p<0.05, η_p^2=0.181. In the first block, there was larger TA in the competitor items than the non-competitor items (marginal by subjects), F1(1,82)=3.701, p=0.058, η_p^2=0.043, F2(1,20)=6.489, p<0.05, η_p^2=0.245. In the second block, there was no difference between two competitor types.

A significant interaction was also found among the competitor, the association, and the cue conditions in the by-subject analysis, F1(1,82)=4.194, p<0.05, η_p^2=0.049, F2(1,20)=1.722, p=0.204, η_p^2=0.079. Further planned analyses were conducted in the human and the physical condition separately. In the human cue condition, no simple interaction between the association and competitor types, but there was significant interaction between competitor and blocks, F1(1,38)=8.362, p<0.01, η_p^2=0.180, F2(1,10)=6.209, p<0.05, η_p^2=0.383. A larger TA score was found in the competitor items than the non-competitor condition in the first block, F1(1,38)=11.104, $p<0.01$, $\eta_p^2=0.226$, F2(1,10)=8.468, $p<0.05$, $\eta_p^2=0.459$, but no difference was found in the second block. In the physical cue condition no simple effect and simple interactions was found (Fs < 1).

First Time Window (200-500ms)

In the first time window (200-500 ms), no main effect was found, but there was a significant interaction between competitor and blocks, F1(1,82)=10.009, $p<0.01$, $\eta_p^2=0.109$, F2(1,20)=7.179, $p<0.05$, $\eta_p^2=0.264$, an interaction among competitor, cue types and blocks, F1(1,82)=4.869, $p<0.05$, $\eta_p^2=0.056$, F2(1,20)=4.741, $p<0.05$, $\eta_p^2=0.192$, and an interaction among competitor, association and blocks in the by subject analysis, F1(1,82)=4.676, $p<0.05$, $\eta_p^2=0.054$, F2(1,20)=0.911, $p=0.351$ $\eta_p^2=0.044$. Further analyses were conducted in the first and the second block separately. In the first block, there was larger TA in the competitor items than the non-competitor items, F1(1,41)=6.518, $p<0.05$, η_p^2 =0.137, F2(1,10)=7.315, $p<0.05$, $\eta_p^2=0.422$. In the second block, there was an interaction between competitor types and the association types in the by-subject analysis, F1(1,41)=6.695, $p<0.05$, $\eta_p^2=0.140$, F2(1,10)=1.645, $p=0.229$, η_p^2 =0.141. Simple effect analysis showed that in the homogeneous condition, there was no difference between two competitor types, but in the heterogeneous condition, there was larger TA in the non-competitor condition than the competitor condition in the by-subject analysis, F1(1,41)=9.046, $p<0.01$, η_p^2=0.181, F2(1,10)=3.948, $p=0.075$, $\eta_p^2=0.283$. In the competitor condition, there was a marginally larger TA in the homogeneous condition than the heterogeneous condition in the subjects analysis only, F1(1,41)=3.720, $p=0.061$, $\eta_p^2=0.083$, F2(1,10)=1.104, $p=0.318$, $\eta_p^2=0.099$. In the non-competitor there was no difference, F1(1,41)=1.210, $p=0.278$, $\eta_p^2=0.029$, F2(1,10)=1.009,, $p=0.319$, $\eta_p^2=0.099$.

In the Human group, there was a Competitor-by-Block interaction, $F(1,38)=12.304$, $p=0.001$, $\eta_p^2=0.245$, $F2(1,10)=8.568$, $p<0.05$, $\eta_p^2=0.461$. Follow up analyses revealed that in the Competitor condition, Target Advantage was greater in the first block than the second, $F(1,38)=9.038$, $p<0.01$, $\eta_p^2=0.192$, $F2(1,10)=3.174$, $p=0.105$, $\eta_p^2=0.241$; in the Non-competitor condition, things were reversed: greater TA in the second block than the first, $F(1,38)=5.037$, $p<0.05$, $\eta_p^2=0.117$, $F2(1,10)=5.344$, $p<0.05$, $\eta_p^2=0.348$. In the Physical-cue group, no main effects or interactions were found, (All Fs < 1.08, ps > 0.30).

Before discussing the results from these two time windows or later windows, we should note that many results are significant by-subjects only. Normally in psycholinguistic research, one takes care to establish significance both in an analysis that treats subjects as a random variable and one that treats items as a random variable because our generalisations are across both. By-items analyses ensure that a significant effect by-subjects is not due to only some of the linguistic materials presented to participants having an effect. Thus we should be cautious about drawing strong conclusions from these results. That being said, we should note that our items differed very little among each other in this study. It may be that terms for shapes may differ in frequency to those for fruits or animals and this could be one cause of the difference. However, there is a more likely cause of variation among items given that the study was designed to induce memory effects. Since items were presented in a fixed order for all participants, it could be that an item in a given condition that is presented early in the study yields a different eye-gaze pattern to one in the same condition presented later in the study, when the effects of learning arise or strategies change. So, although we should be cautious about some of the results to be discussed, we at least have some reason to be a little more confident about by-subject only significance than we might otherwise be in a psycholinguistic study, as long as we understand how things change in the course of a session.

Discussion of results from Preview and 200-500ms:

Turning now to the results from the preview and 200-500ms windows, we can say that these bear out somewhat the observations made from visual inspection of the graphs. In the first half of the study in these windows, there was a higher target advantage overall in the Competitor condition compared to the Non-competitor condition. This is not something that one normally expects in Perspective-Taking tasks. But it should be pointed out that this is not a standard perspective-taking task in that 18 of 96 trials refer to private grid positions. This fact would generally encourage participants to attend to the private objects as well as common objects in the process of a trial. As discussed above, the effect could be accounted for in terms of two kinds of strategy: one of using the presence of two objects to predict the target; a second strategy of using the absence of a second object to initially anticipate a private-position target. Obviously, this account of the pattern found is speculative and calls for further experimental investigation. One possible piece of indirect evidence for the account lies in the fact that the pattern is most apparent in the first block, whereas the results from the second block (to be discussed below) are quite different. If one or both of these strategies are born of a more or less conscious impression participants gained from early trials, this may have worn off as the study went on.

In considering the graphs for the data above, it was commented that the strong ‘reverse’ pattern found in the early time window was more apparent in the Human-cue group than the Physical-cue group. The analysis of the Competitor-by-Cue-by-Block interaction lends support to this observation. Only in the Human group was there more Target Advantage for the Competitor trials in the First Block.

In the second block of the trials, the interaction in the 200-500ms time window revealed no difference in target advantage between Competitor and Non-competitor conditions in the Homogeneous condition but a greater target advantage in the Non-competitor than the Competitor, on Heterogeneous trials. This is the memory-based effect predicted by Horton and colleagues. Returning to consider the second time block results in Figure 2.6, we can see that the difference between Competitor and Non-Competitor TA scores in the 200-500ms widow is more marked in the Human cue group than the Physical-cue group. However, this apparent difference is not directly supported by any statistical results. One small, indirect piece of evidence that more is going on in the Human cue group comes if we look at the breakdown of the interaction among Competitor, Cue-type and Block conditions by Human- and Physical-cue. This shows that there were no significant effects or interactions in the Physical group, whereas the Competitor-by-Blocks interaction in the Human group revealed higher Target Advantage for Competitor trials in the first block but higher TA for Non-competitor trials in the second block. So, it seems that in the 200-500ms window, there is a memory-based effect emerging in the second half of the experiment. It also seems that this may be driven more by the eye-gaze results of the Human-cue group.

500-800ms Window:

In the second time window (500-800 ms), there was a main effect in the competitor types (marginal by items), F1(1,82)=7.734, $p<0.01$, η_p^2=0.086, F2(1,20)=3.940, $p=0.061$, η_p^2=0.162, due to the target advantage score being larger in the non-competitor condition than the competitor condition. There was also a significant interaction among competitor, association and cue types, F1(1,82)=15.404, $p<0.001$, η_p^2=0.158, F2(1,22)=6.050, $p<0.05$, η_p^2=0.232,. Simple effect analyses were conducted in the human and physical condition separately.

In the human cue group, there was interaction between competitor types and association in the by-subject analysis, F1(1,38)=7.972, $p<0.01$, η_p^2=0.173, F2(1,10)=2.100, $p=0.178$, η_p^2=0.174. In the homogeneous condition no difference was found between the non-competitor and competitor conditions, but in the heterogeneous condition participants had larger advantage scores in the non-competitor condition than in the competitor condition, F1(1,38)=8.678, $p<0.01$, η_p^2=0.186, F2(1,10)=5.201, $p<0.05$, η_p^2=0.342. In the competitor condition no difference was found, but in the non-competitor condition, there were larger TA scores in the heterogeneous condition than the homogeneous condition in the by-subject analysis, F1(1,38)=7.931, $p<0.01$, η_p^2=0.173, F2(1,10)=3.123, p=0.108, η_p^2=0.238.

In the physical cue condition there was also an interaction between competitor and association condition, F1(1,44)=7.542, $p<0.01$, η_p^2=0.146, F2(1,10)=6.023, $p<0.05$, η_p^2=0.376. Follow-up analyses revealed that in the homogeneous condition, participants had larger advantage scores in the non-competitor condition than in the competitor condition in the by-subject analysis, F1(1,44)=12.063, $p<0.01$, η_p^2=0.215, F2(1,10)=3.347, p=0.097, η_p^2=0.251, but no difference was found in the heterogeneous condition. In the non-competitor condition there were larger TA scores in the homogeneous condition than the heterogeneous condition in the by-subject analysis, F(1,44)=7.970, $p<0.01$, η_p^2=0.153, F2(1,10)=0.545, p=0.477, η_p^2=0.052, but no different was found in the competitor condition.

There was also an interaction between the association and blocks in the by-subject analysis, F1(1,82)=4.700, $p<0.05$. η_p^2 =0.054, F2(1,20)=1.440, p=0.244, η_p^2=0.067. In the homogeneous condition, there was larger TA in the second block than the first block in the by-subject analysis, F1(1,82)=5.119, $p<0.05$. η_p^2=0.059, F2(1,20)=3.060, p=0.096, η_p^2=0.133, but there was no difference in the heterogeneous condition. In both first and second block, there was no difference between the homogeneous and the heterogeneous condition.

Summary of results from 500-800ms window:

The first thing to notice about the results in the 500-800ms window compared to earlier windows is that overall, bias in Non-competitor trials was greater than Competitor trials. This is the ‘normal’ pattern of results for this kind of Perspective-Taking procedure and it is the opposite of the 'reverse' pattern found in earlier time windows.

Next, it is notable that there was only one effect involving Blocks in this time window, showing that on Homogeneous trials there was greater Target Advantage in the second block than the first, but there was no difference in Heterogeneous trials. This result is generally in line with Memory-based predictions if we work on the assumption that participants should get better at the task as the session progresses. If, as Memory-based accounts predict, the Heterogeneous condition engenders more attention to the private positions, it would explain why improved performance over time only occurs for the Homogeneous group. Again, we can return to the Second-Block graphs in Figure 2.6 to compare the Heterogeneous trial results in the Human and Physical cue conditions for this time window. We see that the difference between Non-Competitor and Competitor conditions looks more marked in the Human Cue group than the Physical Cue group. However, this difference did not manifest itself directly in any statistically reliable effect.

There were however differences between Physical and Human-cue conditions when we break down the three-way interaction in this time window. In the both the Human-cue and Physical-cue groups there were Competitor-by-Association interactions. However, when analysed further they manifest opposite patterns. In the Human group, we can view the fact that Target Advantage for the Heterogeneous condition was lower in the Competitor trials, while there was not TA difference in the Homogeneous condition as yet another manifestation of a Memory-based effect. Things are less straightforward though if we compare Target Advantage on Competitor trials between Homogeneous and Heterogeneous. These were not different, though one might expect them to be if there were memory-based effects. By contrast, a comparison of TA on Non-competitor trials reveals a larger bias in the Heterogeneous condition.

It would be unwise to read too much into these results until further testing, or a better designed study were able to clarify the extent to which the clearly more strategic behaviour from the first half of the experiment affects behaviour in each of the Homogeneous and Heterogeneous conditions in the second half. However, returning to Figure 2.7 it seems that in the second block of trials, the pattern of results in the Human condition does seem consistent with a memory-based effect..

The follow up analysis of the Competitor-by-Association interaction in the Physical-cue group yields a very different picture. Here TA in the Non-competitor trials was greater in the Homogeneous than the Heterogeneous condition, meaning that a TA difference between Non-Competitor and Competitor trials was found only in the Homogeneous condition. One tentative hypothesis about this result could be that participants were trying to use the physical cue as a means to generate expectations about the target. Because these are not as, 'natural' or 'salient' as speaker-based cues, greater memory load was required. When this extra load combines with the complexity of the Heterogeneous instruction set, it leads to a slower response in terms of identifying the target.

4.3. Discussion

In this experiment, a participant had 96 trials. 18 of those trials referred to private-view grid positions. Other things equal, this fact might create in participants some expectation that the target will be in a private grid position and encourage greater attention to those positions. The design also made it possible for participants to associate one of two cues (a specific speaker or a specific physical cue) with those 18 private-view trials. One unforeseen effect of the experimental design seems to have been that it engendered strategies on the part of the participants, particularly in the first half of their experimental session. Our results suggest that these strategies were more prevalent in the Human-cue group. This fact in itself shows a difference in the effect of the cue-type.

To the extent that the second-block results were less affected by strategic activity, these revealed that in the 200-500ms window there was a memory-based effect of the kind one would expect given previous research in Horton & Gerrig (2005a), Horton (2007) and elsewhere. Moreover, there was indirect evidence in these results that the memory-based effect was more pronounced in the Human-cue group.

More evidence for memory-based effects were found in the later time window. In addition, there was a three-way interaction that revealed clearly different patterns of behaviour in the Human-cue and Physical-cue groups. The follow-on analyses produced results that are not entirely clear. One interpretation offered was that in the Human-cue group, something like the predicted memory-based effect emerged in the second half of the study. By contrast, memory-based effects in the Physical-cue group may have been distorted by the fact that the cue used was more difficult to memorise.

Taken together, these results point to the conclusion that memory-based effects can be produced in an experimental procedure that is based on Perspective-taking task. More importantly, given the aims of the study, we have some evidence that not all memory-based cues are equal. Why it is that identity of interlocutor is a better cue than simple association between a single interlocutor and colour/sound is not answered here. Memory-based research has focused solely on Human-based cues. Indeed, Brown-Schmidt et al. (2015) sees interlocutor identity as, “a particularly potent source of contextual constraint” compared to other contextual cues (ibid, p.61). This could be because an individual’s identity may be very easy to process and code in memory, given the frequency with humans identify one another and the importance of other-agent identity to our daily lives. However, things may not be so straightforward since there are very many facets of identity that can be coded and later used as a cue. It is a long-established fact in social psychological research that that people use gender, race, professional status and categories like university major or birthplace to keep track of information in conversation (Taylor & Falcone, 1982; van Twuyer & van Knippenberg, 1995; Frable & Bem, 1985). One conclusion from this social psychological research is that participants tend to code in memory those aspects of another agent’s identity that are relevant in the prior context.

Even if facets of another agent’s identity are in some sense easier to code in memory than other contextual cues, there could be other factors that make agent-related information more susceptible to cued retrieval at later stages. In Chapter 1, I outlined proposals about how the social cues involved in conversational interactions trigger processing biases that would integrate situational information more deeply for relevance. A by-product of this process would be that information shared with another agent is integrated more thoroughly in memory and is thus more susceptible to cued retrieval at a later point. In Experiment 1, the stimuli in the physical cue condition were salient for the participant alone. They were not in shared attention for participant and speaker. The speaker gave instructions by video link and did not acknowledge them or show any sign of awareness of them; and indeed the confederate speakers were completely unaware of these stimuli. By contrast, in our social-cue condition and previous studies supporting a memory-based account, the cue, being the speaker alone, was in shared attention at the coding and retrieval stage. To see this, recall Barwise’s description of shared attention to an utterance event as involving attention to the situation in which both agents attend to that event. Thus the speaker was in shared attention when the previous utterance was made.

Experiment 2 would provide a further test of the domain-general memory-based hypothesis and my own, socially constrained memory-based hypothesis. That study was designed to explore to what extent memory-based processes were constrained to encode and retrieve information previously in shared attention.

5. Experiment 2

The previous study provided evidence for the memory-based model, and demonstrated that participants had less interference from their privileged knowledge when the speaker had a stronger association with the common-ground referents. But this cue-based process seems not unconstrained because when physical cues were matched with the referent participants were less liable to establish any association, even though the salience of the physical cues seems rich enough. An explanation proposed by the memory-based model suggested that the association is only successfully made when the memories associated the highly salient cues are ready, that is, the stronger the cue is, the easier the association between the speaker and the referent is to make. The partner-specific association will not influence the subsequent language processing if it is too weak to reach the threshold (Horton & Gerrig, 2005a). However the last study indicated that how to define the strength of the cues is not only based on the perceptual salience, but also in terms of the likelihood that the cue receives attention and is integrated into memory together with relevant information. My proposal about the role of previous shared experience in common ground suggests that social factors are critical in any memory-based maintenance of common ground.

If joint attention influences the employment of the low-level cues in the memory-based mechanism as I conjecture in the last experiment, we should expect to find greater competitor interference to the extent that participants jointly attend to the cues for the homogeneous and heterogeneous conditions. As in the first experiment the present study also used the type-based and location-based descriptions to refer to objects in the grid. Two association conditions were operationalized through the different distribution of these two types of instructions. Like the last experiment, the homogeneous condition has all the instructions referring only to objects in common ground, including the location-based filler instructions. The heterogeneous condition has three quarters of location-based fillers making reference to privileged ground objects, so that the speaker gives instructions about both common ground and privileged ground objects. In each session, the participant receives the full set of homogeneous and heterogeneous instructions in a random order. Based on the results of the last experiment, we can predict that more competitor interference would be found in the heterogeneous condition than in the homogeneous condition, to the extent that the participant responds to the different cues for each instruction set.

In the present study both conditions used a physical cue to each of the homogeneous and to the heterogeneous set of instructions. The physical cues were simplified as only colours instead of colours and sounds, so it would be easier for participants to encode their memory. I.e. participants saw one colour (e.g. red) when they were hearing an instruction from the homogeneous set and another colour (e.g. blue) when hearing an instruction from the heterogeneous set. In order to make the colour cue more prominent, we coloured the private grid positions rather than the border of the image of the speaker. See Figure 2.8.

To test the hypothesis that joint attention influences the memory of the association between the speaker and the referent, a variable regarding the shared attention to cues was manipulated in a between-groups design. For both groups, participants had to click on the name of the colour of the background appearing with a given trial, from a choice of two (blue and red) on the centre of their screen (see Figure 2.9). In a shared cue condition, the speaker also asked what colour is the participant's side, and repeated the answer after the participant replied. In a non-shared cue the choices were presented with a text question on the screen only. The participants viewed the question and made the choice all by themselves, so that the speaker never showed his attention of his partner's privileged information. Thus, in the shared-cue condition, on each trial, a colour cue to homogeneous or heterogeneous instructions was in shared attention. In the non-shared cue condition, they were not.

According to the memory-based model, the present study should show the effect of the association as the first experiment (homo- vs. heterogeneous instructions). What's more, if memory-based model is constrained by the shared knowledge, there should also be an effect of the shared attention. Specifically, in the shared cue condition, participants should get less interference from their privilege ground in the homogeneous condition. But in a non-shared cue condition, the association effect should be weaken, because an associated memory representation could not be established when the information was not jointly attended. However, if the fact that participants cannot build up the memory of association under the physical cue in the first experiment is simply due to the property or salience of the cue itself, the manipulation of the shared attention should not make any difference. According to this prediction, the physical cue itself does not vary between conditions, so there should be no difference between both the shared and non-shared cue conditions.

5.1. Method

5.1.1. Participants

40 participants (10 male and 10 female in the shared cue condition; 11 male and 9 female in the non-shared cue condition) were recruited from UCL's participant pool. One UCL student from the Department of Linguistics was recruited to play the speaker role. This recruit was blind to purpose of the study. All were paid for their participation.

5.1.2. Stimuli and design

As in Experiment 1, participants were given instructions to move objects around a 3*3 grid. The set of objects in each trial was chosen from the same pictures as in Experiment 1 from the three categories: fruit, animal or geometric shape. To reduce the possibility that participants attended to the common ground object which had the same subtype competitor in their privileged ground, all the trials, except the non-competitor condition, had two objects with the same subtype, one in the common ground and one in the privileged ground. All the instructions were pre-recorded by the same person, thus each participant in a given set of conditions heard the same instructions. The instructions consisted of type-based and location-based instructions as Experiment 1.

There were 20 critical trials and 20 filler items in each association condition. As in the first experiment, 20 critical trials consisted of 10 competitor trials and 10 non-competitor distractor trials, and they all involved type-based instructions. The 20 filler items involved location-based instructions. In a homogeneous association condition, all 40 instructions were targeted at objects in common ground. In a heterogeneous condition, the 20 critical instructions and 5 of the 20 location-based filler instructions were targeted at the common ground objects, but 15 of the 20 filler instructions were targeted at privileged ground objects.

In order to lessen the presence of strategic effects found in Experiment 1, the proportion of filler trials containing dual objects was changed. In the first experiment 14 out of 48 fillers had dual objects. Now, all location-based filler trials had dual objects. This means that a participant would have a total of 60 dual-object trials. On 20 of these trials (the Competitor experimental trials) the target would be the common-ground token of the dual object (e.g. the commonly viewable apple). On 40 dual object trials, the target would another one of the objects. On 15 of the latter

40, the target would in fact be a private-view object. Thus, 1/3 dual object trials has the common dual object as target. Given that there are three common object, this represents a chance distribution.

In order to distinguish the association conditions, the participants' side of the shelves had two colours. Half the instructions had the participants' side of the shelves in red, while the other half had the side in blue. See Figure 2.8. One colour accompanied homogeneous instructions while the other accompanied the heterogeneous instructions. These were counterbalanced across the experiment.

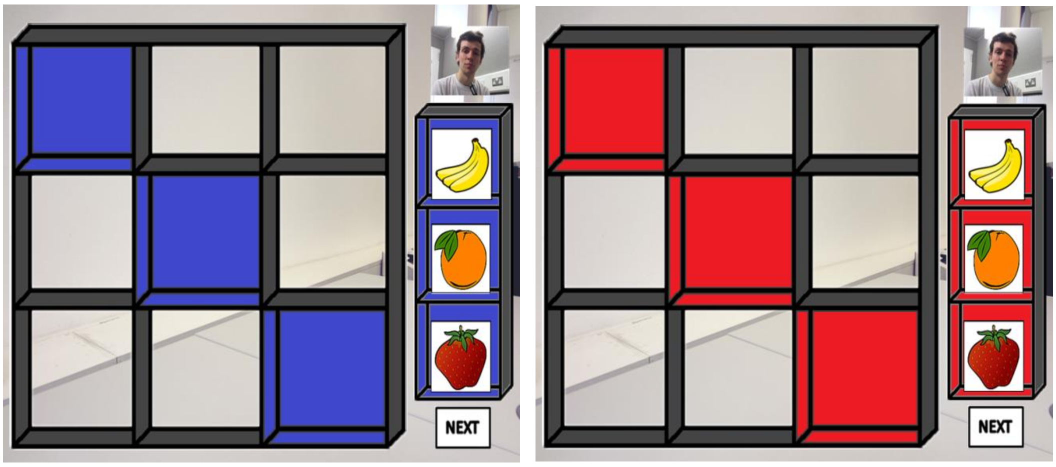


Figure 2. 8. The initial display in Experiment 2.

Two colours accompany a Homogeneous or a Heterogeneous condition respectively. Participants were presented both colours.

The shared-attention condition was operationalized in a between groups design. Participants were presented a question about the colour on the screen before the speaker gave instructions. In a shared-cue condition, the speaker asked what colour is the participants' side, then repeated the answer after participants clicked on the choices. In the non-shared cue condition participants received the question and chose the answer on the screen silently, while the speaker remained ignorant of the answer.

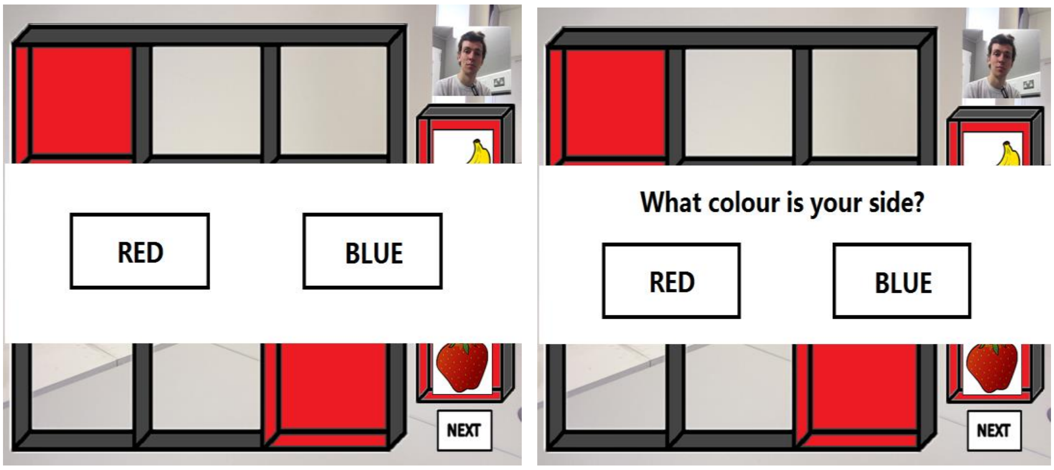
To sum up, the experiment had a 2 (competitor) * 2 (association) * 2 (shared attention) design. Altogether, each session with a participant involved 80 trials, 40 delivered in the homogeneous and 40 in the heterogeneous condition. In each association condition, 10 trials referred to the non-competitor targets and 10 referred to the competitor targets. Two types of critical instructions were matched and used the same recordings. As was the case in Experiment 1, the trial order was fixed across both groups. It was constructed using a randomised order except that three location-based fillers were placed among the first ten trials, prior to the occurrence of the first heterogeneous test trial.

Shared attention was a between subjects manipulation. So a participant either selected the answer after the speaker's question (shared cue condition) or the screen prompt (non-shared cue condition).

5.1.3. Procedure

In each session, participants were seated in front of a 23” screen fitted with Tobii TX300 eye-tracking equipment in a room with only the experimenter. They were told that they would be interacting with a person located in another part of the college via a video link. They were then shown the same initial display at in Experiment 1.

The training phase was the same as in Experiment 1 with both a participant and the speaker taking an alternate role of an addressee or an instructor. Firstly, the 3*3 grid appeared on the left of the screen. On the right there was a shelf with three objects which were not visible to the speaker. The participant was required to move the objects to the coloured grids of the right shelf, and then three common objects were presented. After that a new step was added on the original procedure. Participants were presented a question, "What colour is your side?", either from the speaker or on the screen. They were required to click the answers on the prompt in both conditions. See Figure 2.9. Finally participant moved a target object according to the speaker's instruction and clicked the ‘NEXT’ button on the bottom right of the screen. After every two trials, participants had an opportunity to break prior to re-calibration of the eye-tracking equipment.



a. Shared attention condition b. Non-shared condition

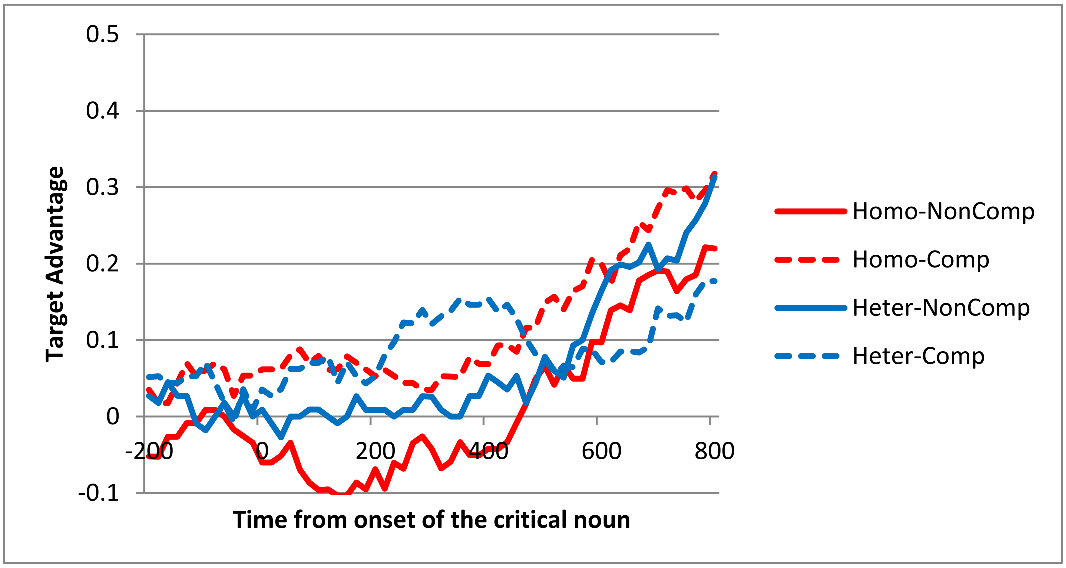
Figure 2. 9. The question sessions in both conditions in Experiment 2.

When the speaker asked the colour verbally, participants were only presented the answers (Shared attention condition. see the left.). In order to indicate that the speaker shared the knowledge of the colour, he also repeated the answer after they made choice. When the question were shown silently, only participants knew the colour because the speaker would not receive and repeated their answer (Non-shared condition, see the right.)

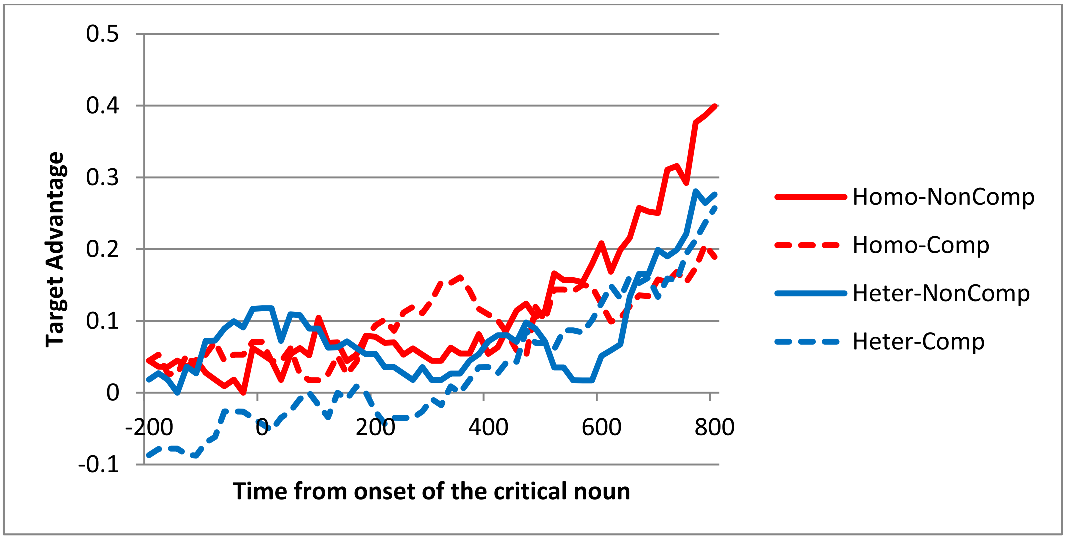
5.2. Results

The same eye-movement analysis was conducted as in Experiment 1. The target and competitor items were identified on a by-trial basis. Target advantage score, ln(P(target)/P(competitor/non-competitor)), was analysed for three time regions from the onset of the critical word: a preview region (-200ms to 200ms), the first time window (200ms to 500ms), and the second time window (500ms to 800ms). As the study investigates memory-based effects, Blocks are again included as a factor.

Figure 2.10 shows a plot of the time course of target advantage for the two groups. Figure 2.11 shows the average target advantage score for each of the competitor and non-competitor conditions for each of the three windows. And Figure 2.12 shows the proportion of fixations to the target and the competitor in eight conditions.



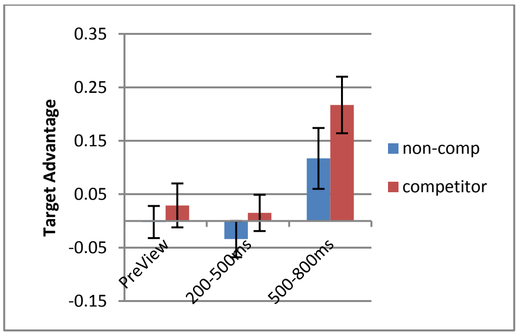
Shared cue condition



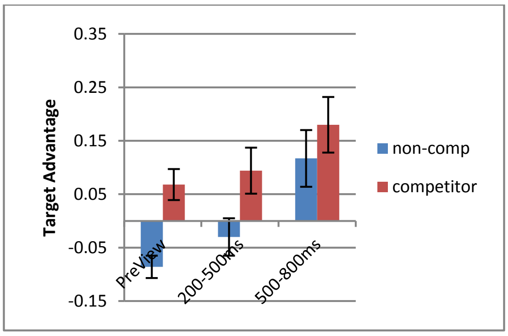
Non-shared cue condition

Figure 2. 10. The time course of average target advantage score in Experiment 2.

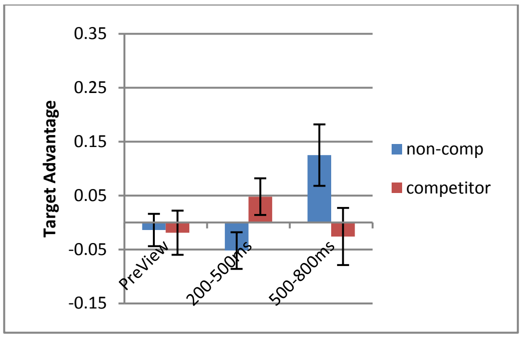
It shows the TA scores in the human cue condition (upper) and physical cue condition (lower) from the 200 ms before the onset of the critical word to 800 ms after the onset.

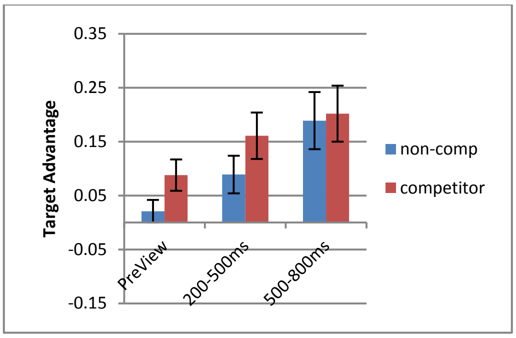


First block - Homogeneous



Second block – Homogeneous

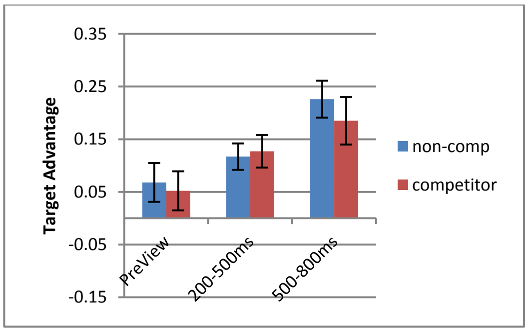


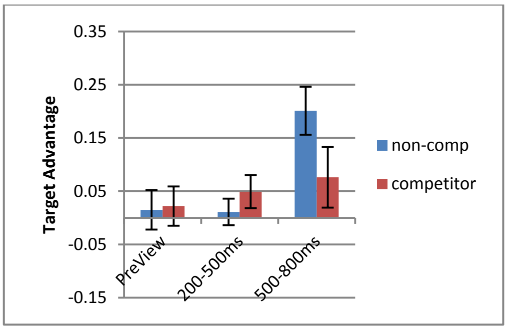


First block - Heterogeneous

Second block - Heterogeneous

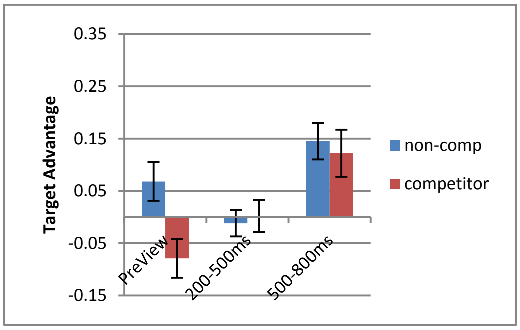
a. Shared condition

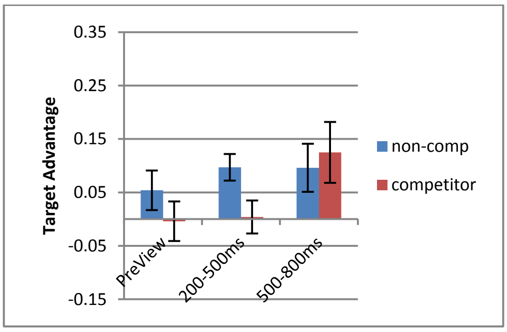




First block - Homogeneous

Second block – Homogeneous





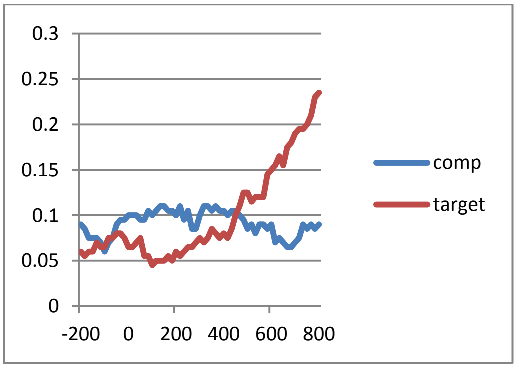
First block - Heterogeneous

Second block - Heterogeneous

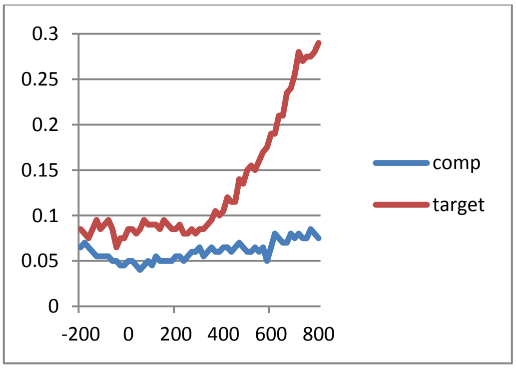
b. Non-shared condition

Figure 2. 11. The average target advantage score for the three time windows in Experiment 2 by block.

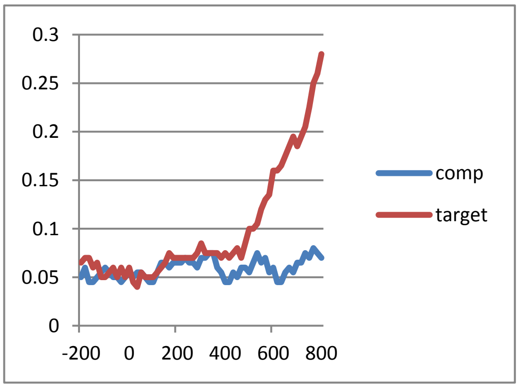
The charts show the competitor and non-competitor conditions in the shared cue (2.11a) and the Non-shared cue (2.11b) conditions. In each case, the averages are shown for the first block (left) and second block (right) for the homogeneous (upper) and heterogeneous (lower) conditions. Error bars indicate standard error of by-participant means.



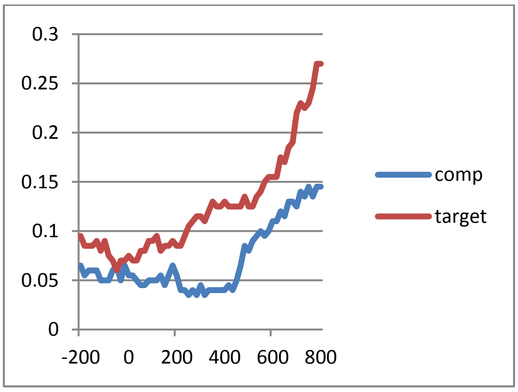
Homogeneous – Non-competitor



Homogeneous – Competitor

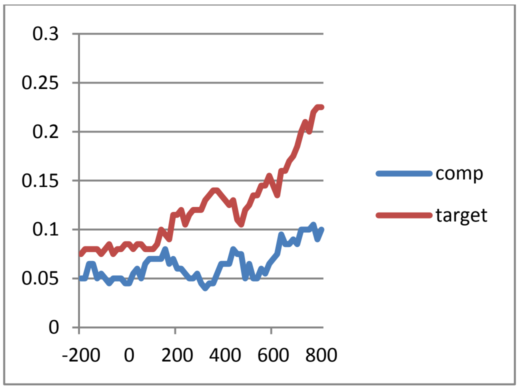


Heterogeneous – Non-competitor

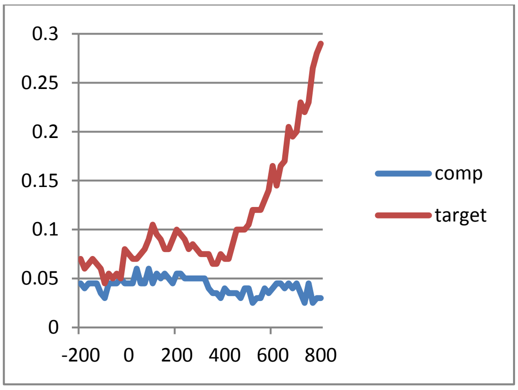


Heterogeneous– Competitor

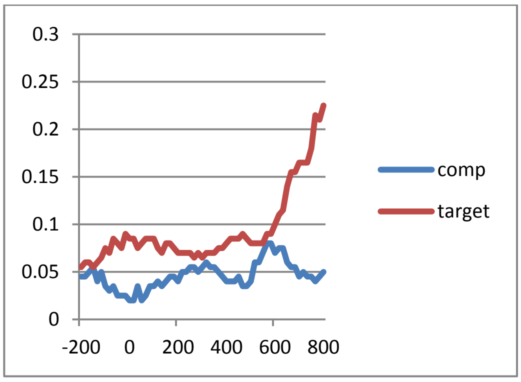
a. Shared condition



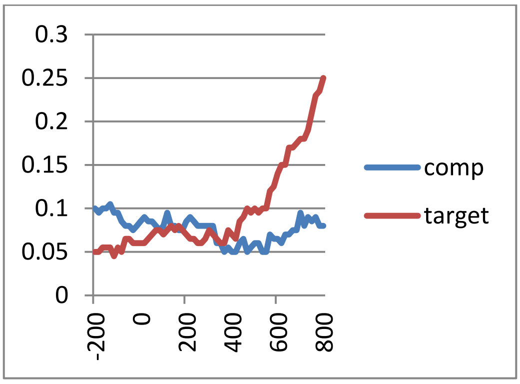
Homogeneous – Non-competitor



Homogeneous – Competitor



Heterogeneous – Non-competitor



Heterogeneous– Competitor

b. Non-shared cue condition

Figure 2. 11. The proportion of looks to the target and the competitor in Experiment 2.

They show the homogeneous-noncompetitor, the homogenous-competitor, the heterogeneous-noncomeptitor, the heterogeneous-competitor conditions in either the shared (first four) or the non-shared (last four) conditions.

Before turning to consider the statistical analyses let us consider the graphs in these figures. The first thing to note, in the context of Experiment 1 is that there looks to be a ‘reverse’ pattern in

Target-Advantage scores again, particularly in the Shared-cue group, Homogeneous condition. Recall that in Experiment 1, the reverse pattern was strongest in the Human cue, Homogeneous condition. There is a difference to Experiment 1 in that the ‘reverse’ pattern seems stronger here in the second block of trials, although it is somewhat present in the first block also, but only for the Shared-Homogeneous condition. It was noted in Section 5.1.2 above that the counterbalancing for Experiment 2 was changed from how it was in Experiment 1. Here we included proportionally more dual-object trials (i.e. trials where the display had two apples, one in common and one in private view). The rationale was to make the chances one in three that when a dual object display occurred, the target would be the common-view token of that pair (i.e. the target in Competitor conditions). Given that there are three common-view objects, the idea was that dual-object displays would not be a cue to the target in Competitor trials. It seems that this change in the balance of visual displays has been counter-productive. Either the strategies participants developed in Experiment 1 have been reinforced by this change or the change has led to a new strategy, manifesting itself in the second half of the experiment.

One speculation that can be offered here is that some participants developed a ‘process of elimination’ strategy. Symeonidou et al. (2015) report on a perspective-taking task similar to that employed in Keysar et al. (2000) where participants first looked at object that they were not going to choose before forming a bias to the target that they eventually choose. If participants adopt this strategy here, it would be to the disadvantage of Non-competitor trials since it would take longer to establish that there is not a dual object in private view, than to find the second object when it is there. If we assume that participants are looking for a dual object unsuccessfully in Non-Competitor trials, it would explain why there is a negative or zero bias in these trials where the ‘reverse’ pattern is manifest – i.e. in the Shared Homogeneous and Heterogeneous trials.

At this stage, we can only offer speculations as to what explains 'reverse' patterns. For the time being, we turn to the statistical results.

5.2.1. ANOVA analyses

A mixed 2(Competitor vs. Non-Competitor)*2(Homogeneous vs. Heterogeneous)*2(Shared vs. Non-Shared)*2(First Half vs. Second Half) ANOVA revealed the following results.

Preview window:

In the preview window, there were no main effects. There was an interaction between the competitor types and the cue types, F1(1,76)=5.008, $p<0.05$, η_p^2=0.062, F2(1,16)=6.359, $p<0.05$, η_p^2=0.284. Simple effect showed that for the non-competitor items there was larger TA in the non-shared cue condition than the shared cue condition in the by-subject analysis, F1(1,76)=5.907, $p<0.05$, η_p^2=0.072, F2(1,16)=3.752, p=0.071, η_p^2=0.190. But in the competitor items there was no different TA score between two cue types.

200-500ms Window:

In the first time window (200-500 ms), no main effects were found. There was a two-way interaction between association and the cue in the by-subject analysis, F1(1,76)=4.764, $p<0.05$, η_p^2=0.059, F2(1,16)=2.584, p=0.128, η_p^2 =0.139. But follow up analyses revealed no effects or interactions. There was also an interaction between the association and blocks (marginal in the by-items analysis), F1(1,76)=5.943, $p<0.05$, η_p^2 =0.073, F2(1,16)=4.195, p=0.057, η_p^2 =0.208. Follow up analyses showed in the heterogeneous condition, there was larger TA in the second blocks than the first blocks, F1(1,76)=6.672, $p<0.05$, η_p^2 =0.081, F2(1,16)=9.351, $p<0.01$, η_p^2=0.369, but in the homogeneous condition, there was no different TA between two blocks, F1(1,76)=0.331, p=0.567, η_p^2=0.004, F2(1,16)=0.482, p=0.497, η_p^2=0.029. In the two blocks there was no difference between two association types (Fs < 3.17, ps >0.1).

Summary of results in 200-500ms window:

In these time windows, there was only one significant effect that involved Blocks as a condition. This showed only that Target Advantage increased in the Heterogeneous condition from the first to the second block but no difference between the first and second half was found in the Homogeneous condition. This is in some ways the opposite of a memory-based effect. If we assume that performance improves over the course of the experiment, we would expect Target Advantage in the Homogeneous condition to increase more than in the Heterogeneous, if an association between Heterogeneous trials and private-view targets is made. It may be that this effect results from the fact that in the Shared Heterogeneous condition, the 'reverse' strategy discussed above only develops in the second half of the experiment, leading to better performance on Competitor trials.

500-800ms window:

In the second time window (500-800ms), there were no main effects. A three-way interaction among competitor type, association and cue types was found, $F1(1,76)=5.413$, $p<0.05$, $\eta_p^2=0.066$, $F2(1,16)=8.823$, $p<0.01$, $\eta_p^2=0.355$. Follow-up analyses showed that in the shared-cue condition there was no simple effect but a significant simple interaction between competitor type and association type, $F1(1,38)=5.238$, $p<0.05$, $\eta_p^2=0.121$, $F2(1,8)=5.764$, $p<0.05$, $\eta_p^2=0.419$. Simple effect analysis showed that in the non-competitor condition there was no difference between the heterogeneous and the homogeneous condition, $F1(1,38)=0.564$, $p=0.457$, $\eta_p^2=0.015$, $F2(1,8)=0.023$, $p=0.883$, $\eta_p^2=0.003$, but in the competitor condition a larger TA score in the homogeneous condition than the heterogeneous condition, $F1(1,38)=4.500$, $p<0.05$, $\eta_p^2=0.106$, $F2(1,8)=9.839$, $p<0.05$, $\eta_p^2=0.552$. In both the homogeneous and heterogeneous condition, there was no difference between the competitor and non-competitor conditions.

In the non-shared cue condition no main effect or interaction was found (Fs < 3.2, ps > 0.1).

Summary of results in 500-800ms window:

On the face of it, the results in this window confirm the predictions made here in the thesis. On competitor trials in the Shared group, Target Advantage is higher in the Homogeneous than Heterogeneous, while there is no difference in the Non-Competitor condition. Additionally, no such interaction is in evidence in the Non-shared cue group. However, we should be cautious about these results in the context of the apparent strategy that seems to be used, particularly in the Shared group. Still, if whatever strategy is at play both in the Homogeneous and Heterogeneous trials, one could argue that the eye movements have been modulated in the Heterogeneous trials by attraction of the Private-view competitor object. If we consider the graphs in Figure 2.12, showing proportion of looks to target and competitor objects separately, we can see that in the Shared, Heterogeneous Competitor condition, a strong bias forms to the competitor object, in a way that it does not in the corresponding Homogeneous condition. This contrast is not apparent in the Non-shared graphs.[17] Thus, there is some evidence that the memory-based effect is more prominent here in the Shared group than the Non-Shared group.

Finally, although there were no significant effects in the Non-Shared group, a 2*2*2 ANOVA (excluding Blocks as a factor) reported in the first submission of this thesis did show a significant two-way interaction also in the Non-shared group when the three-way interaction in this 500-800ms window was followed up. As in Experiment 1, follow up analyses on this interaction showed the opposite effect compared to the shared group, where a difference in Target Advantage between Non-competitor and Competitor only occurred in the Homogeneous condition. As with the first experiment, we speculate on this result that there

[17] This impression is backed up by statistical analysis that takes looks to the competitor as the DV. This is reported in Zheng & Breheny (under review).

is a memory effect here but it is weaker due to the greater memory load, perhaps because it arises from a more conscious strategy.

5.3. Discussion

Although not entirely conclusive, the results indicated that in the shared cue condition participants demonstrated a memory-based effect. In particular, during the period 500-800ms after the actual word onset participants demonstrated larger interference from their privileged distracter when the speaker had previously referred to participants' privileged ground. Since the average length of the critical word was around 600ms, this time period corresponds only to the point where participants will have processed the critical word (allowing 200ms as the time between making a decision and launching an eye movement).

The results suggested that a greater association in memory was created between the physical background colour and the patterns of instructions (homogeneous/heterogeneous) and this association influenced the subsequent referential processing. It is noteworthy that the cue effect was greatly enhanced when the speaker created shared attention to this cue through a question.

6. General discussion

The memory-based model proposed an automatic domain-general association process for prioritising previously shared information in conversation (Horton & Gerrig, 2005a, Horton, 2007). The two experiments reported above demonstrate the operation of a memory-based mechanism in the context of a perspective-taking task. Participants received two sets of instructions associated with two different cues. One set of instructions frequently made reference to private grid positions. We found that, for some cues, participants suffered more interference from privately viewable objects when the cue was to the heterogeneous instructions, compared to when the cue was to the homogeneous instructions. It seems then that the association between the cue and the past references to privileged grid positions led to more attention to these positions and thus more interference from competitor items held there. This is arguably an effect of memory mechanisms of the kind described in the literature as being responsible for making previously shared information salient.

However, in both experiments, the increased interference effect was modulated by the cue. In both experiments, we found one type of cue to be more effective than another. For the first experiment, we compared the effect of speaker identity as cue with speaker-plus-physical cue and found that speaker identity was more effective. This difference may have been due either to simple salience of the respective cues, or, as I argued, to the fact that cues are more effective when in shared attention. The second experiment directly tested the hypothesis that cue needs to be in shared attention at the coding phase. We found the predicted increase in interference when the cue was in shared attention. When the cue was privately attended by the participant, we found no memory-based effect.

On reflection, it may seem obvious that memory-based processes could not operate unconstrained and still properly serve the function of supporting the processing of referents in common ground. After all, one could be interacting with a speaker whilst attending to private thoughts, or one could simply be co-present with another agent while attending singly to an object or event, and in neither case would the object of attention be common ground. If memory-based mechanisms indiscriminately reflect memories of past events in which the current speaker was co-present, then the extent to which their output aligns with true common-ground information would be much diminished. Our results show that memory-based mechanisms do not operate unconstrained, and seem to reflect a selection from what is in shared attention at the point of encoding. The results thus support my proposals in two ways. First, we have established that decision processes about referents are affected by low-level memory based processes. Our position is similar to that of Horton and colleagues, that these mechanisms can provide rapid, implicit and relatively low-cost access to previously shared information. But unlike other memory-based models, it seems that information previously in shared attention is prioritised. This latter result provides support to a second element to my proposal – that information in shared situations, what is shared attention, receives differentially more attention and is processed more thoroughly for relevance and so is integrated more deeply with context.

Casting our mind back to the review of research on toddlers we found that the operation of a common-ground mechanism can sometimes work against the optimal functioning of decision processes for definite reference (cf. Moll et al. 2011). When the cues to shared information were strong, children seemed unable to treat an unseen object as novel for an adult experimenter, in spite of the fact that the adult clearly could not see the object. Something similar has happened in the studies I report in this chapter. We developed a procedure wherein cues to common ground result in increasing the salience or activation of objects that the director cannot see. Thus, in perspective-taking tasks, where a participant needs to consider what the director knows about objects in privileged ground, it could be that mechanisms for promoting common ground objects work against referential decision processes. This is a possibility that we will explore in the next set of experiments. Before that, we will consider in detail previous studies of perspective-taking.

Chapter 3: Perspective-taking on-line

- What on-line research tells us about the use of common ground and Theory of Mind in referential processing.

In this chapter we will consider research that applies time-sensitive visual-world methods to investigate referential processes during what have become known as referential game or perspective-taking tasks. Perspective-taking tasks have grown out of a long line of research on interaction in dialogue (Krauss & Weinheimer, 1966; Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987) where interlocutors' knowledge and beliefs diverge from one another but where they have to co-ordinate on some facts in order to perform the task. In perspective-taking tasks, participants are typically faced with a situation where the speaker is not able to see a set of objects that they can see themselves but where there are other objects that are jointly visible.

1. The Perspective-taking Task - Keysar, Barr, Balin, & Brauner (2000)

An important early study of this sort is reported in Keysar et al. (2000). Participants were presented a grid of 4 by 4 slots containing objects and required to reorganize these objects around the grid with the help of a director (an experimental assistant) (see Figure 3.1). Some slots were only visible to participants, while others were visible to both sides. The director instructed participants to move objects in the task. In a test condition, two unequal sized objects (e.g., a large and a medium candle) were presented in shared slots so that both director and participant could see them, and a third, competitor object with a different size (e.g., a small candle) was presented in a participant privileged slot - only visible to participants. In a control condition, the participant privileged competitor (the small candle) was replaced by an irrelevant object different to the shared objects (not a candle), so both participant and director were presented only two contrastive objects. Participants' reactions were coded and analyzed after hearing utterances containing the critical phrase “the small candle”. In the test condition, the referent of this phrase was not the smallest one from the participant’s perspective, although it was the smallest one from the director’s view. Thus participants had to ignore the fact that a visible object best fit the description and consider how the director viewed the grid in order to determine the correct reference. Results showed that participants had more fixations and a longer looking time to the participant-privileged competitors in the test condition compared to the irrelevant objects in the same privileged positions in the control condition. Participants also reached and grabbed the privileged competitor more in the test condition compared to the irrelevant objects in the control condition. These results indicate that participants consider the privileged competitors as referents in the test condition. Besides, the looking time latency showed that participants had faster initial eye movement to the privileged competitor than the shared target in the test condition. And both their initial and final fixation on the shared target in the test condition was delayed compared to the control condition. These results indicate that when a private competitor existed, participants took it into consideration initially.

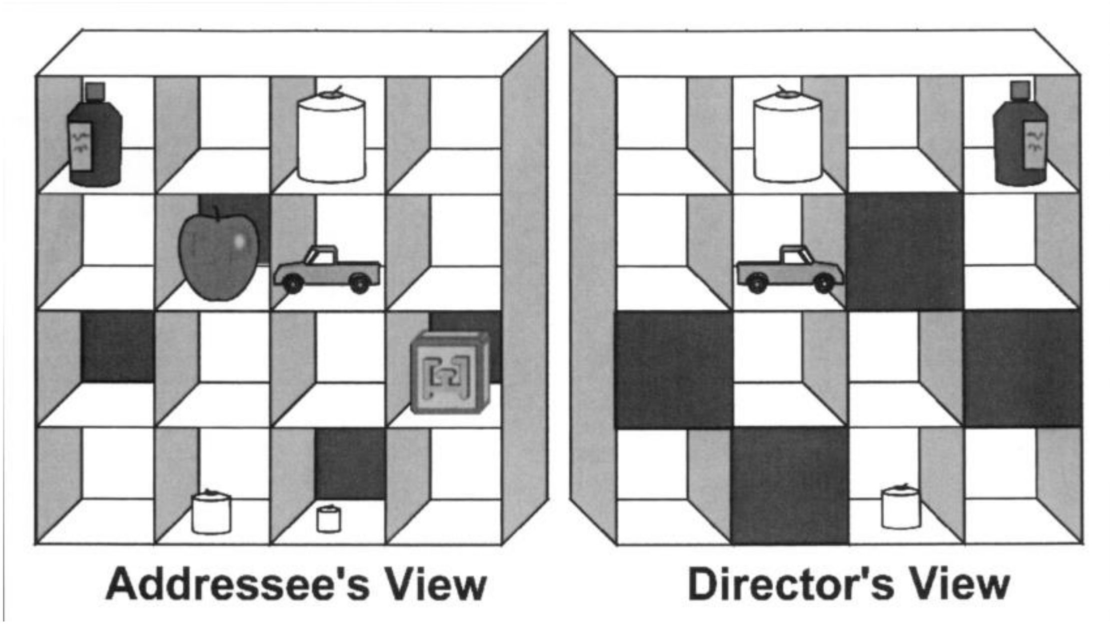


Figure 3. 1. The sample setup in Keysar et al. 2000's study.

Addressee and director were sitting in each side of the shelf. The white slots were visible to both the addressee and the director, but the black slots were only visible to the addressee. (Keysar et al. 2000, p. 33.)

2. Perspective-Adjustment Model

As Keysar et al. argue, if the commonly viewable objects are considered to be in common ground and if the privately viewable objects are not, then the results of this study challenge the theory of referential processing found in Clark & Marshall 1981 and elsewhere, which says that definite referents are constrained to be common ground. Participants clearly find it difficult to ignore the object which is the best fit for the description, 'the small candle', in so-called privileged ground.

The results of the study also provide support for Keysar et al.'s proposed, 'perspective adjustment' theory according to which comprehenders initially process referents from an egocentric perspective, only to later adjust where information about the speaker's perspective is integrated into the decision process. The perspective adjustment theory is motivated by the idea that keeping track of what the speaker knows and does not know is a cognitively costly process and, as such, makes the strategy of ignoring that perspective temporarily worthwhile since one's egocentric perspective and the correct, speaker-adjusted perspective are frequently aligned.

The Perspective-adjustment model first proposed by Keysar and colleagues assumes two operations: an unrestricted search and a monitoring process (Keysar, Barr & Balin, 1998; Keysar, Barr & Horton, 1998; Keysar, 2007). Through the former operation the addressee searches potential referents unrestrictedly, while this search process is monitored by an adjusting process with higher-level and meta-knowledge memory structures. This monitoring role is attributed to common ground by Keysar and colleagues (Keysar, Barr & Balin, 1998; Keysar, Barr & Horton, 1998; Epley, Morewedge, & Keysar 2004; Keysar, 2007).

3. The role of common ground mechanisms in perspective-taking tasks

In setting out the operation of the perspective adjustment model, Keysar and colleagues typically assign the role of the costly higher-level mechanisms to be that of monitoring what should be common ground. However, as discussed in previous chapters, keeping track of common ground information does not involve making inferences about the speaker's mental state for this kind of study. In particular, the common assumption among researchers who use the perspective-taking paradigm is that the commonly visible objects in these studies are established as being common ground in virtue of their co-presence. In terms of Clark & Marshall's proposals, this means that a co-presence heuristic is applied to the scenario to determine the common object as being in common ground.

This heuristic does not call on higher-order inferences, according to Clark & Marshall. As mentioned above, the co-presence heuristic requires evidence for shared attention to co-present objects. Specifically, the heuristic could not apply to objects that are not commonly visible to rule them out from common ground. There are two reasons for this. The first is that it would be entirely inefficient to continuously monitor an utterance situation for objects that are not jointly visible so as to rule these out of the common ground. Common ground is a positive construct only. As a constraint on definite reference it would contain a pool of possible referents for interlocutors to refer to with definite forms. It is incoherent to conceive of common ground processes yielding information about what is not in common ground.

The second reason for a co-presence heuristic to not exclude objects from common ground would be that objects can enter common ground via other means, such as through prior discourse. Just because an object is not co-present, it does not mean that it could not be common ground.

In terms of the proposal made in Chapter 1 about common ground, I would similarly argue that keeping track of common ground via cues to joint or shared attention would be relatively low in cost.

So, if success on this perspective-taking task is difficult due to the interference that the privately viewable objects create, then this is not due to any difficulty maintaining common ground, by any theory of common ground, but more likely to due to other factors. One such factor is that the best match for the description that the speaker uses (‘the small candle’) is to the privileged object. Another is that the procedure adopted in Keysar at al.’s experiment does not ground the commonly viewable objects sufficiently well. For example, by simply viewing the objects, people may not share their attention on the target objects. A third is that, in this context it may be difficult to integrate information about the speaker’s ignorance of the third candle in the on-line decision process about the referent. Indeed, the application of theory of mind abilities in this context seem to be critical to success, even though these abilities are largely inert here in keeping a record of common ground.[18]

One further point to consider about the experimental procedure in Keysar et al. (2000) and many subsequent studies using similar procedures is that it is not entirely clear that the privately viewable objects are not in common ground, at least according to the theory of common ground being proposed in this thesis. To see this, consider that at the beginning of a session, the participant and director are told about the set up of the grid, involving private slots for the participant and critically, both director and participant are instructed that there will be objects in the grid visible only to the participant. Thus, via verbal instruction, it becomes common ground that there will be objects in the participant's private slots. So, it is common ground that there are objects there when the director gives instructions. What is not common ground, of course, is the identity of the objects (whether there are any candles there, for instance). A strict application of Clark's theory would mean that these privately viewable objects are not in common ground under the description used by the director ('the small candle', for instance) and are thus not available for definite reference under that description. However, a theory of common ground that does not serve as a necessary pre-condition for definite reference but is a construct that promotes potential referents to the extent that they have been previously shared, would say that even privately viewable objects are common ground to some extent, perhaps less than the commonly viewable objects. We will take up this point further below.

To sum up this discussion, we have reviewed an influential study which seems to show that common knowledge of definite referents (under

[18] Recall the observation made in the first chapter that it is impossible to apply ToM reasoning abilities to establish a referent as common knowledge since it would involve an infinite series of every higher-order inferences about the other agent's higher-order inferences. Clark & Marshall are aware of this problem and that is why they propose heuristic processes for inferring common knowledge.

the description provided) is not a hard constraint on referential processes. These results add indirect support to the conceptual arguments put forward in Chapter 1 that, from a cognitive perspective, it is not necessary that a target definite referent's identity be common knowledge. We have however taken issue with the idea that an egocentric heuristic procedure is motivated by the cost of maintaining information on what is common ground.

4. Graded Salience Models and the Cost of Perspective-Taking

Stepping back from the specifics of the perspective-taking task, even when considered in terms laid out by Clark and colleagues, common ground is not necessarily costly to maintain. One could perhaps argue that a diary representation of previous encounters requires extra resources over and above those needed to access encyclopedic memory for discourse processing.[19] For example, one could claim that diary representations require an explicit form of episodic memory of previous encounters. However, no argument for such a requirement exists and it is difficult to conceive of why explicit memory would be essential to maintain common ground. Instead, it seems sensible to allow that Clark & Marshall's heuristics could be implemented via implicit memory mechanisms, of the kind assumed in memory-based models.

Brown-Schmidt (2012) in fact proposes something along these lines. That paper demonstrates a graded effect of cues to common ground. In one condition, an experimenter asks about the identity of an object privately viewable for the participant. At test, this context elicited greater visual bias to that object as a potential referent than in the case where the

[19] Keysar et al., (1998) seem to make such a suggestion but it is unclear what they have in mind as being required to maintain common ground, as they make frequent reference to 'higher-order' processes. This suggests that, like many researchers in this field, they merge the functioning of theory of mind in perspective-taking tasks with the maintenance of a diary representation for common ground.

participant was prompted to announce the identity of the object to the director. This result seems echoic of the result in Moll & Tomasello (2007), reviewed above, that showed different degrees of interaction yielded different degrees of awareness of sharedness at test. In another paradigm, Metzing & Brennan (2003) demonstrate a partner-specific association between a referring term and target referent. Metzing & Brennan's account of their results is cast in terms of Clark & Marshall's theory but assuming that associations between speaker and referring form are made in memory, becoming available later via automatic cue-based activation.

Thus, there is an idea abroad that maintaining common ground, even as envisaged by Clark & Marshall, could be integrated into normal memory processes for discourse interpretation in general. However, this idea is found together with the idea that success in perspective-taking tasks is achieved simply by constraining referents to common ground. Here we have taken issue with this idea. To exclude a privately viewable object as a possible referent requires more than mechanisms for maintaining common ground. It requires an inference about the speaker's ignorance of the identity of the private object. That is, it requires a Theory of Mind inference.

To return to the question of costs involved in perspective-taking tasks, we are left with the question why participants have difficulty with the task presented in Keysar et al. (2000). Above we mentioned three kinds of factor that are commonly discussed in the literature. Two of these are specific to the design of Keysar et al.'s task and are addressed in subsequent literature, to be reviewed in coming sections. The third factor that is considered concerns the cost of theory of mind inferences. We will also review literature that bears on whether access to ToM inferences on-line are necessarily costly. We shall see that subsequent research in this area finds an early effect of perspective-taking, calling into question the specific of Keysar et al.'s proposal. Finally, we identified a fourth possible factor in this section that is yet to be explored experimentally. This is the possibility that, to some extent, the privately viewable objects are common ground, in the sense that their existence is shared information, even though their specific identities are not. If social cognitive mechanisms for discourse generally promote the activation of potential referents that are shared, then it could be that these mechanisms work against the participant in this task. We will explore this idea in the next set of experiments, presented in the following chapter.

5. Constraint-based research

In language processing research, the constraint-based model proposes that all relevant and available information, including both linguistic and non-linguistic information, is continuously integrated into language processing (e.g., McRae, Spivey-Knowlton, & Tanenhaus, 1998; MacDonald, Pearlmutter, & Seidenberg, 1994; MacDonald, 1994). The theory is also extended to information that is relevant in perspective-taking tasks. In this way, perspective information is just one of the general constraints on language processing. All the constraints compete with each other, and the one with stronger weight will overcome others to determine the favoured interpretation.

According to proponents of the constraint-based model, ground information and perspective information are processed in parallel with linguistic information and other contextual information, so it can be employed without delay in language processing (Hanna et al., 2003; Brown-Schmidt, & Hanna, 2011). Numerous studies have demonstrated eye gaze patterns which suggest that perspective information is taken into account early. In Hanna and colleagues' 2003 first experiment, they explored how definite reference was disambiguated through an incremental language process (see Figure 3.2). In the task participants attempted to put a shape (e.g. a triangle) on another particular destination shape which was depicted by a definite noun phrase such as “the red one”. Two identical destination shapes were always present at the same time, with a target shape being in the common ground and a competitor shape being either in the common ground or addressee's privileged ground. The colour of two destination shapes were also manipulated as same vs. different. Accordingly, the utterance would be ambiguous only when the same colour competitor was in the common ground (the left upper chart in Figure 3.2), rather than the different colour competitor (the lower charts in Figure 3.2) or the same colour one in addressee's privileged ground (the right upper chart in Figure 3.2).

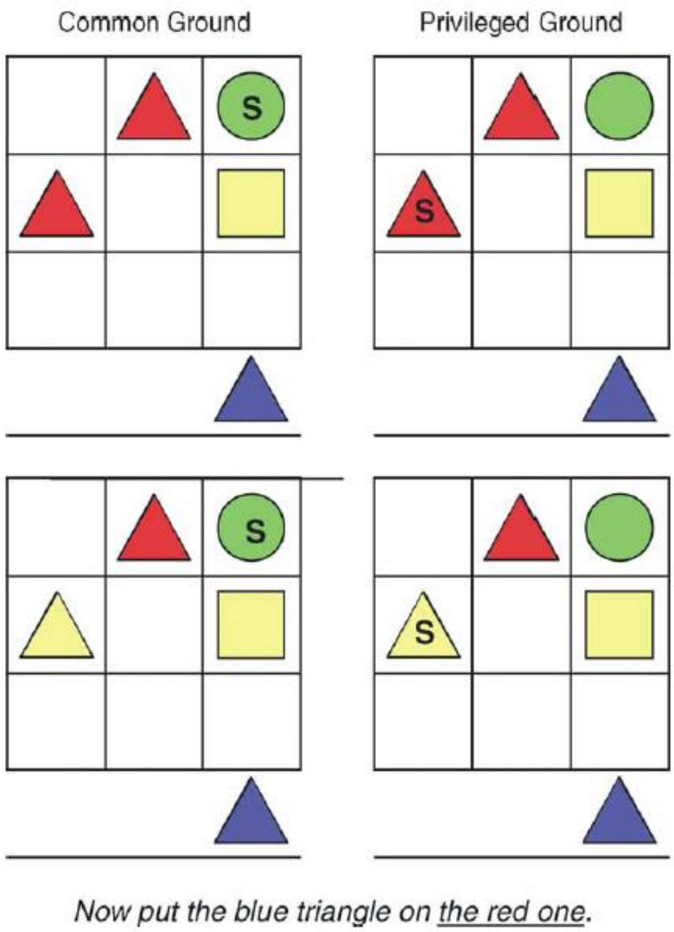


Figure 3. 2. Sample displays in Hanna et al., 2003’s study.

Participants were required to move the blue triangle on a destination triangle which had a competitor with a same/different colour in the privileged/common ground. The left upper display illustrates the same colour common ground condition, the right upper is the same colour privileged ground condition, the left lower is the different colour common ground condition, the right lower is the different colour privileged ground condition (Hanna et al., 2003, p.46).

The results showed that when the competitor was in the different colour, the looking time to the competitor was not different to any other irrelevant objects. When the same colour competitor was in the common ground, a bias in eye gaze formed equally on both the target or the competitor. A critical point was, when the same colour competitor was in addressee's privileged ground, whether the looking pattern would be more analogous to the same colour/common ground condition, or the different colour condition. Results showed that participants mostly looked at the target initially, and the looks rose steadily and quickly, which was not different to the different colour condition. So it indicated that perspective information was integrated in the same time course as other information. However, similar to Keysar and colleagues' experiment results (see Keysar, Barr & Balin, 1998, Keysar et al., 2000), this study also replicated the finding that privileged ground somewhat interfered with referent search. When the same colour competitor was in addressee's privileged ground, although previous target advantage analysis revealed the same pattern as the different colour condition, a competitor fixation analysis showed the same manner as the same colour/common ground condition, that is, participants looked at the competitor more and longer than any other irrelevant objects. In short, the analysis on the target revealed an early use of speaker’s perspective, but the analysis on the addressee-privileged competitor shows a continued attraction to these referents.

Notwithstanding the continued attraction of the privately viewable object, these results directly challenge the proposals of the perspective-adjustment model. According to that model, perspective information is integrated into on-line processes at a secondary phase, preceded by the deployment of an egocentric heuristic that equally considers all potential referents from the participant's perspective. Hanna et al. report results that show a clear early sensitivity to perspective.

Constraint-based models blend Hanna et al.'s two contrasting findings by proposing that different weighted constraints compete with each other to compute an optimal interpretation on-line. In particular, compositional language processes involved in these tasks require comprehenders to identify a contrast set (a set of candles in Keysar et al.'s study, a set of triangles in Hanna's study) due to the modification of the description. For instance, 'the small candle' makes implicit reference to a contrast set of candles. Comprehenders need to find a referent for this set. It is reasonable to assume that there is a strong bias to assign all candles in the visual array as the referent of this implicit variable, in the absence of strongly constraining factors. Keysar et al.'s argument could be that perspective information is just such a constraining factor and should block the private candle from membership of this set. However, the argument from the constraint-based camp would be that the set-up of Keysar et al.'s study does not make perspective information sufficiently salient or relevant to the task. The procedure in Hanna et al. (2003) involved a set up phase where director and participant jointly established the common-ground objects via a verbal process. In addition, the participant was given their one, 'secret' object to place where they wished. In this way, the procedure reinforced perspective information, providing more shared attention to the common-ground objects, but more attention to the speaker's ignorance of the private object. Thus, the difference in terms of time-course of the use of perspective information between the two studies could be attributed to procedure.

Another factor, mentioned above, has to do with the fact that in Keysar et al.'s study, the description, 'the small candle' best fits the private object, while in Hanna et al.'s study, both private and common objects are equally good fits to the description. But this aspect of Hanna et al.'s design has led to a criticism of its value to show the normal time course of deployment of perspective information. For instance Barr (2008) notes that in Hanna et al.'s design, there is one condition where there are two red triangles in common ground. This procedure creates a global ambiguity in a number of trials. As global ambiguities do not get resolved without further recourse to the speaker (e.g. by asking), participants might be prompted to be generally more strategic in the task, employing otherwise costly perspective-taking abilities from the outset. Thus a follow-up study eliminating this feature, reported in Heller et al (2008) is relevant. Here, the design is such that there is never a global ambiguity, only a temporary one. The pattern of gaze data reported in Heller et al. is similar to that in Hanna et al. Generally, there is a clear early effect that participants integrate perspective information (by excluding the privileged object as a member of a contrast set for a modifier), but there is also a greater proportion of looks to the privileged competitor compared to a non-relevant distractor in privileged view. Heller et al. interpret this latter finding somewhat differently to Hanna et al. in that they propose that participants do not exclude all attention to private objects in pursuit of a common-ground-only heuristic, but monitor objects in private as well as common ground. Heller et al. cite a study in Wardlaw Lane et al. (2006) showing that speakers can be over informative in their descriptions, integrating private objects into the contrast set for a modified description (i.e. they describe a single triangle in common view as a 'the small triangle' when there is a larger triangle in private view). However, it is unclear why participants should monitor private objects in performing these tasks. We return to this question below when we consider the idea that privately viewable objects are, to some extent, common ground.

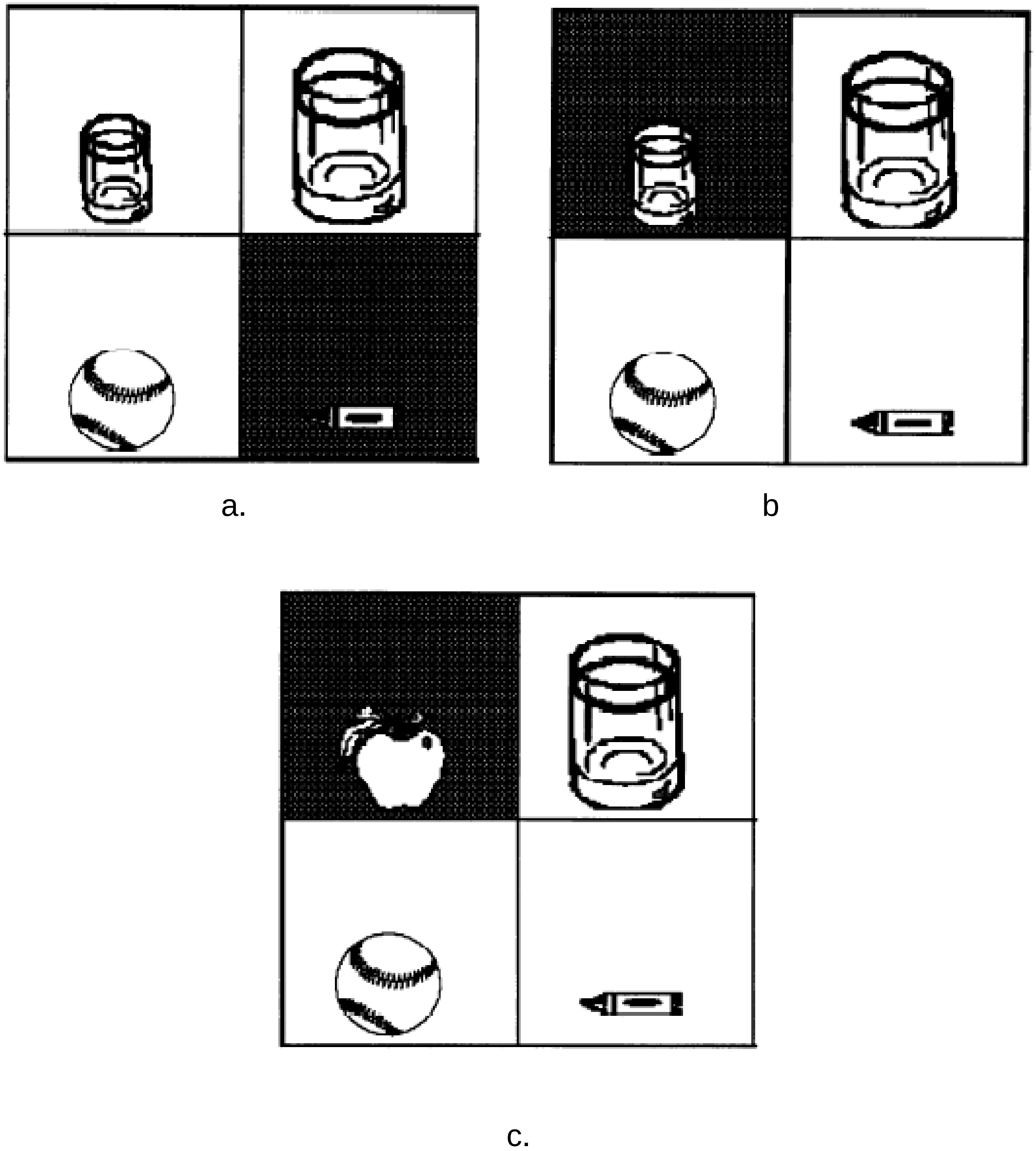


Figure 3. 3. Sample displays in the three conditions in Nadig and Sedivy 2002's experiment.

The common ground condition is displayed the upper left, the privileged ground condition at the upper right, and the baseline condition at the bottom. The write area is shared between the speaker and the addressee, and the black area is only visible in the participant-addressee's view (Nadig & Sedivy, 2002, p.331).

The evidence of early integration of common ground is also provided in a child study. Nadig and Sedivy (2002) used the similar, but smaller, visual display as Keysar and colleagues' 2000 study. They presented to 5- to 6-year-old children an array of 2 by 2 slots with four objects. Two identical objects were presented but with unequal sizes (e.g., a tall and a small glass). In a common ground condition both objects were shared (Figure 3.3.a), whereas in a privileged condition the competitor object was participant-addressee privileged (Figure 3.3.b). A baseline condition was also performed with all four objects were in different kinds (Figure 3.3.c). When a confederate-speaker asked children to pick up the target (e.g. the glass), it is ambiguous in the common ground condition but not in the baseline condition. The critical examination focused on children's behaviour in the privileged condition. A target advantage analysis (the fixation duration to the target (the bigger glass) subtracted the fixation duration to the contrasting (or control) object) (the small glass or the apple) demonstrated a significant difference between the privileged ground condition and the common ground condition, but not between the privileged ground condition and the baseline condition in an early time window from the onset of the critical word. This result indicated that children initially looked at their privileged competitor to the same extent as the privileged control object. Children did not refer to their privileged knowledge despite of the presence of the competitive object. Therefore, as early as five years old children are competent with integrating the common ground in the early stage of comprehension. The experimental display is similar with Keysar and colleagues 2000’s study, and children are supposed to be less competent in perspective taking than adults in that study. But the result contradicts the late monitoring role of common ground as Keysar and colleague’s study suggests. In their study more objects are involved, and the privilege competitor is the best fit of the description (e.g. the small candle). In contrast in this study the display is simpler and the description is equally matched with both shared and privileged objects (e.g. the glass). It indicates that the complexity of information and the burden for memory resource may influence the effect of common ground. The constraint-based model can better explain the results in both studies, because this model predicts different degree of deployment of perspective information based on the different competition between the perspective constraints and other language constraints.

6. A role for mental-state information in perspective-taking tasks?

In previous sections, we have taken care to distinguish between mechanisms for maintaining common-ground information, via various heuristic processes, and mechanisms for inferring information about the speaker's individual perspective (i.e. their mental state) and integrating that information into the referential decision process. The latter mechanisms would deploy theory of mind abilities in on-line processing. In perspective-taking studies where it is shown that participants do not simply include all visible objects in their initial search, such as reported in Hanna et al. (2003) and Heller et al. (2008), it is unclear whether this is achieved by the deployment of a strong common-ground constraint only, or whether information about what could not plausibly be a referent also plays a role. To clarify this question, consider that in the constraint-based (C-B) model, common ground is a strong constraint on referential processes. In probabilistic terms, this means that whatever referents are common ground are promoted as highly likely referents. According C-B models, the more cues there are to common ground, the greater the likelihood that common ground objects are referents. Likewise, the greater the linguistic cues are to a referent, the greater the likelihood that is assigned to that object as candidate referent. We could thus consider the results of the studies reviewed thus far in this chapter as demonstrating the trade-off between constraints created by common ground and constraints created by bottom-up linguistic processing. It may be that information about the speaker's perspective, what they know about and are ignorant of plays a peripheral role, or no role at all - at least not initially.

There are alternatives however. These would see a greater role for actual perspective information, in the form of what the speaker does not know about, in referential decision processes.

According to my proposals about the nature and function of common ground in on-line language processing, common ground mechanisms are insufficient to exclude the competitor object in private view. Information from Theory of Mind inferences would be needed for successful, early target bias in these tasks. Recall that my proposal about common ground is that it is a mechanism that promotes the salience of information previously shared with the interlocutor. In common with the C-B model, the degree to which this mechanism plays a role in perspective-taking tasks depends on the degree to which objects are processed in shared attention. But there are significant points of contrast between my proposals and current proponents of the C-B model, all of whom endorse Clark & Marshall's theory of definite reference, for whom common ground is the main constraint.

The first point of contrast is that, contrary to Clark & Marshall's theory, we do not believe that referential processes are strictly constrained to choose among common-ground objects. The standard perspective-taking paradigm developed in Keysar et al. (2000) and employed frequently since is unable to determine whether this is the case since the target is always in common ground prior to the utterance. But in Chapter 1, we discussed a number of examples of accommodation where the referent is not in common ground prior to the utterance and where the audience is required to make inferences about the speaker's mental state in order to determine reference. One such example involved Jane placing mozzarella cheese in the fridge and her flatmate later placing cheddar cheese in the fridge. When Jane later asks for 'the cheese', Sue infers that Jane is referring to the mozzarella, even though the mozzarella is not common ground prior to the utterance being made and even though there are two cheeses in the fridge. To get the right referent, Sue has to make inferences about Jane's false belief about the contents of the fridge (that there is just one kind of cheese there). Such examples differ from the set-up of the standard perspective-taking experiment since they combine the phenomenon of so-called accommodation into common ground with the presence of a competitor that the speaker is ignorant of.

As alluded to in Chapter 1, there is a tradition in pragmatics in line with the approach founded by Grice, that sees referential processing as intimately involving inferences about the speaker's mental state in order to determine their informative intentions (see Sperber & Wilson, 1986; Thomason, 1990 i.a.). According to this tradition, Theory of Mind inferences should be integrated routinely into referential processes.[20] The approach to definite reference adopted in this thesis is aligned within this tradition. By contrast, Clark (see Clark, 1996) relies more on external cues to common knowledge of discourse-relevant facts. Apart from the common-ground heuristics discussed above, research in this tradition has focused on the role of features of discourse such as grounding signals ('uh-huh') and signals for lack of grounding ('huh?'). It is consistent with this tradition that Theory of Mind inferences are peripheral to definite referential processes, playing perhaps a monitoring role along the lines envisaged by Keysar and colleagues.

The second point of contrast lies in the way the common-ground status of an object is viewed. In Clark & Marshall's theory, an object is in common ground for a given definite description, 'the F', only if it is common ground that the referent is an F. This is so since Clark & Marshall's common-ground constraint exists to ensure certainty for interlocutors that they have coordinated on the same referent. In my proposal, by contrast, common ground is more a bi-product of the normal functioning of memory mechanisms combined with domain-specific biases in attention triggered by socially relevant cues. Thus a referent would be common ground for an

[20] We also endorsed the alternative to Clark & Marshall's Common Ground Constraint. This was the Identifiability Constraint. This is a cognitive constraint on definite reference that follows from the proposals of Russell (see Neale 1990).

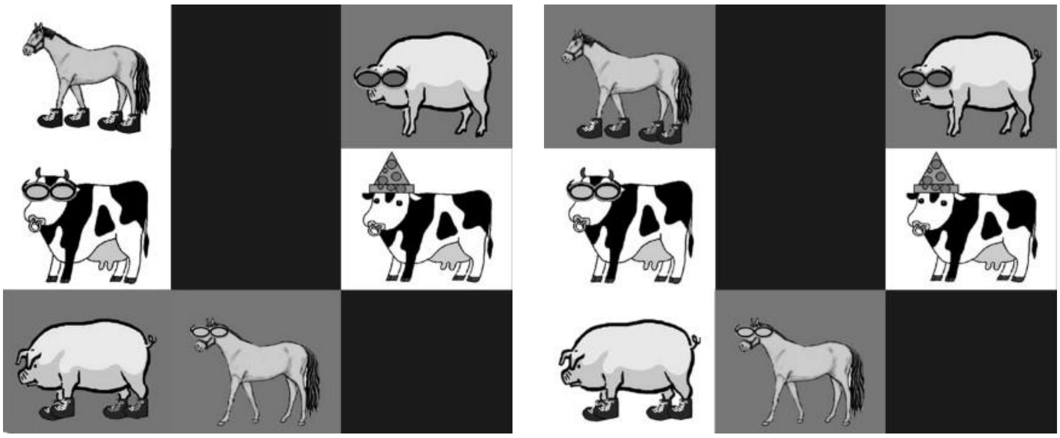
utterance to the extent that it has been processed as an object of shared attention up to the point that the utterance is made. Something which has been shared attention is liable to be salient when an utterance is made by the agent with whom attention was shared, more than an object that has not been shared attention, other things equal. An important consequence of my more mechanistic view of 'common ground' is that an object is in common ground when an utterance of, 'the apple' is made even if its identity as an apple is not common ground. This point is relevant to the observation made above that in perspective-taking tasks the privately viewable objects are in common ground, to some extent. If such is the case, then these objects become salient as referents when the director is giving an instruction. Thus referential processing in Perspective-Taking tasks would rely more heavily on inferences that exclude privately viewable objects (ignorance inferences).

In brief, according to the proposals set out in this thesis, we should expect an important role for Theory of Mind abilities in the standard perspective-taking task. The extent to which participant's bias to the correct target forms in early phases of processing and the degree of competitor interference will depend to a significant degree on the extent to which Theory of Mind inferences, specifically about the speaker's ignorance of the identities of the private objects, are facilitated for the participant. My position contrasts with that of C-B theorists that adopt Clark & Marshall's theory of definite reference. This more widely endorsed approach does not rely heavily on the integration of Theory of Mind inferences in successful completion of the perspective-taking task. Although the C-B view does not rule out the possibility that ToM inferences can support and complement the deployment of the Common Ground constraint, it predicts that success on the task is not reliant on the deployment of ToM inferences.

We turn therefore to consider whether there is any evidence relevant to these differing predictions. Recall that, in the standard versions of the perspective-taking task, of the kind presented in Keysar et al., (2000), Hanna et al., (2003), and Heller et al., (2008), one is not able to tell whether success on this task is conditioned by Theory of Mind inferences since, at least according to Clark's theory as it is adopted in the C-B framework, success could be solely due to the Common-Ground constraint. In the next section, we review what evidence there is for a role for ToM inferences in perspective taking tasks and on-line language processing generally before we move to a general recap of the questions we take forward to the next set of experiments.

7. Evidence for Theory of Mind inferences in on-line language processing

An important piece of evidence that participants integrate information about what the speaker is ignorant of comes in Brown-Schmidt, Gunlogson, & Tanenhaus (2008) (see also Brown-Schmidt, 2009a). The study worked with the idea that if a speaker asks a question about what is in a grid position, they do not know the identity of the object in that position. The design was set up so that, if participants could integrate information about what grid positions the speaker is ignorant about, then they could anticipate the target before a critical point in the question. In a reference game participants were faced with a display that indicated what speaker and themselves could both see, what only they could see and blacked-out grid positions containing objects only the speaker could see (see Figure 3.4.)



a. Early Condition b. Late Condition

Figure 3. 4. Sample displays in Brown-Schmidt et al. 2008's study.

Participants either (a) have one object above the cows in private view and one in common view (Early condition); or (b) both objects above the cows are in private view (late condition). Commonly viewable positions are white, privately viewable positions are grey and blacked-out positions are viewable only to the other player.

There are two conditions. In each condition there are two commonly visible animals of the same species (cows in Figure 3.4). The speaker will ask about the identity of an object that is above (or below) one of these objects. In both conditions the object of inquiry will be privately viewable to the participants – i.e. the speaker is ignorant of that object. In one condition, only one of the animals above these two known animals is unknown to the speaker (early condition – Figure 3.4.a). In a late condition, both animals above the known animals are unknown to the speaker (Figure 3.4.b). When the speaker asks, 'What's above the cow with the glasses?' in the late condition, the participant would be unable to anticipate the target of the question until 'glasses' since both objects above the two cows are unkown to the speaker. In the early condition, however, if the participant can integrate the fact that the speaker asks about the object he does not know, they can anticipate the target after 'cow'. Accordingly, from the earliest time period after the onset of the noun, 'cow', Brown-Schmidt and colleagues found a significantly greater target bias in the early condition. Thus, participants' eye gaze shows clear evidence that they are able to integrate information about speaker ignorance, at least when being asked a question.

It is interesting to note that in discussion, Brown-Schmidt et al. frame their results in terms of integrating ground information.[21] Could it be that participants approach this task relying on ground information only? For example, could participants use a heuristic that maps questions onto information that is not common ground? It seems somewhat implausible that this is the case. It is not true that the target of questions (at least, genuine requests for information) is the same as information that is not common ground. Typically, a genuine request for information is made by someone who is ignorant of that information. Information that does not happen to be common ground does not include only information that one's interlocutor is ignorant of. Much information may be known by two people without that information being common ground. Still, it is perhaps possible that participants in this study adopt a strategy to equate the domain of an inquiry with what is not common ground, as suggested by Brown-Schmidt and colleagues. In that case we need to see if there is any other evidence that Theory of Mind inferences are integrated into on-line language processing.

Ferguson & Breheny (2011, 2012) present a series of studies that show participants anticipate targets by integrating information about either a character's beliefs and desires or about a live speaker's false belief. In neither of these studies is ground relevant to the task. In the case of Ferguson & Breheny (2011), participants here a story about a character who wants to keep a particular desire secret and so acts in a way that would not lead others to think she has this desire. In Ferguson & Breheny (2012), the speaker has an out of date belief about the location of a target object and is answering a question asked by a third party about that object,

[21] See, for example, page 1132.

according to that out of date belief. These studies demonstrate the integration rich Theory of Mind inferences into incremental language processing.

8. Questions Moving Forward

The two main models proposed to account for results in the perspective-taking task both assume that there are automatic processes and corrective, monitoring processes. They differ only in what information can be integrated in the automatic phase. The perspective-adjustment model of Keysar and colleagues proposes that initial processes are from an ego-centric perspective and are not affected by information about what is common ground. These processes are monitored by a system that has access to ground information, among other sources, and can be corrected where necessary. Constraint-Based proposals, found in Hanna et al. (2003), Brown-Schmidt et al. (2008) and elsewhere counter that ground information does get integrated in initial, automatic processes. The idea here is that perspective-taking tasks give rise to a trade-off between common ground constraints and bottom-up constraints that attempt to match linguistic input with targets in the visual context. Corrective processes are necessary where cues to ground information do not sufficiently constrain the set of potential referents to common ground.

It is easy to read the literature on perspective taking and come to the conclusion that Theory of Mind abilities are fundamental to the derivation of ground information. For instance, in a widely cited review of perspective taking literature, Brown-Schmidt & Hanna (2011) say in their introductory section, “In this chapter, we focus on the use of one type of non-linguistic information in incremental language comprehension—information about others’ knowledge and beliefs.” (Brown-Schmidt & Hanna, 2011 p. 12) but the rest of the paper exclusively reviews work that explicitly sets out to test whether common ground information influences incremental referential processing. That review paper also explicitly points to the work of Clark and colleagues as providing the theoretical underpinning for ideas about common ground in referential processing. As discussed at length in Chapter 1 of this thesis, Clark (1992, 1996) views common ground as common knowledge of discourse-relevant information, in line with prominent views in linguistics and philosophy (e.g. Stalnaker 2002). Dating from the work of Clark & Marshall (1981) there is a proper awareness in this work that one cannot establish what is common knowledge in via an infinite series of inferences about ever higher-order beliefs.[22] Clark & Marshall propose that common knowledge has to be established via heuristic means. Thus while it may be true that ground information is information about knowledge or beliefs, because it is common knoweldge that is in question, cognitive processes that establish ground information are not liable to call on Theory of Mind abilities.

Thus, both of the widely discussed sets of proposals about behaviour in perspective-taking tasks are compatible with the idea that Theory of Mind information is only available to be integrated at the monitoring and corrective stage of referential processes. It seems fairly clear that it is the view of Keysar and colleagues that Theory of Mind information is integrated at the secondary stage. For instance, Keysar, Lin and Barr (2003) present a version of the perspective-taking task and explicitly argue that the errors incurred in that task are due to the difficulty in integrating Theory of Mind inferences into language processes. As we have seen above, however, other studies show an effect of early integration of Theory of Mind information in language comprehension (Ferguson & Breheny 2011, 2012). We have also seen apparent evidence in Brown-Schmidt et al. (2008) for integration of ignorance inferences in a taks very similar to the standard perspective-taking task, involving questions. However, Brown-Schmidt and colleagues present their results

[22] Of course, one can establish that something is not common knowledge by an inference that one's interlocutor does not believe some proposition. But the focus here is what is in common ground (hence, what is common knowledge).

as yet more evidence that ground information is integrated on line, extensively citing the work of Clark as providing the theoretical underpinning for their design. This is the case in Brown-Schmidt et al. (2008), Brown-Schmidt (2009), and Brown-Schmidt & Hanna (2011). The proposal seems to be that interrogative forms are processed using ground information - questions only ask about information that is in privileged ground.

It is of course compatible with Constraint-Based views that Theory of Mind inferences, being just another source of contextual knowledge, are integrated into automatic language processes. However, it was argued above it is in the spirit of Clark's work on dialogue that Theory of Mind play a peripheral role, perhaps along the lines envisaged by Keysar and colleagues. This is so since common ground (i.e. common knowledge of discourse-relevant facts) is for Clark the focus of dialogue and common ground is maintained in Clark's work via external cues (see especially Clark & Wilkes-Gibbs, 1986; Clark, 1996). It is really the forms of behaviour that are exploited in the process of moving pieces of the 'dialogue game-board' around.

So, moving forward into the next chapter, we have an open question about the specific role for theory of mind inferences in the perspective-taking task. To date, no research has explicitly studied the role of Theory of Mind inferences in the standard perspective-taking task. More broadly, we are interested in the extent to which processes that maintain common ground play a role in language processing and the extent to which processes that integrate theory of mind inferences play a role. Work in the spirit of Clark sees the lion-share of the constraints coming from common-ground processes. By contrast, my own proposals see a lesser role for common ground and a greater role for Theory of Mind processes. Specifically, regarding the perspective taking task, we have raised a second question about whether common ground processes in fact create some interference in this task.

Chapter 4: The role of common ground and Theory of Mind in perspective taking tasks

1. Common Ground

Throughout this thesis, I have used the term, 'common ground' to refer to what is shared information among interlocutors. I have tried to show that it is possible to take two perspectives on this phenomenon. On the one hand, Clark's perspective and that of many other theorists is that common ground involves common knowledge of discourse relevant facts. For this to be a psychologically plausible perspective on common ground, Clark and others recognise that it is necessary to adopt something like Lewis' theory of common knowledge. According to that theory, it is possible to recognise information as common ground if there is a 'reflexive common indicator' available to the interlocutors. Two key such indicators are the basis of Clark's co-presence and previous-discourse heuristics. On closer inspection, however, it was argued that both of these heuristics rely on cues for shared attention to relevant information. Shared, or joint, attention is a more basic triadic relation than common knowledge. It has been argued in the literature that shared attention provides the basis for shared information in communication (Barwise, 1988; Peacocke, 2005). The alternative perspective, proposed in this thesis, is that common ground is what results from shared attention in current or previous interactions, including previous discourse.

Two further points of contrast between my thesis and the more widely held perspective of Clark's inform the experiments in this chapter. On the one hand, Clark and Marshall's original argument for a role for common knowledge in definite reference had to do with the need for certainty in reference. This necessity results in the common ground constraint on definite reference: definite referents' identities have to be common ground. By contrast, we argue that, from a cognitive processing perspective, mechanisms for common ground would often be unable to provide the right referent of definite expressions due to the phenomenon of accommodation, as shown in one of the cheese examples discussed in Chapter 1. Instead, we motivate the presence of common ground in communicative interactions on the basis of research in communicative development. This research shows that linguists and philosophers' intuitions about the 'openeness' involved in communication (what Tomasello et al. (2005) terms, 'knowing together') likely stem from how infants' concept of communication develops, supported by cognitive biases for processing information in shared situations. These biases yield a greater integration of shared information in memory as such information is processed more deeply for relevance (Csibra & Gergely, 2009).

Thus our alternative perspective on common ground links in with a mechanistic approach involving domain-general memory mechanisms, constrained by social-cognitive cues. Common ground mechanisms then promote the salience of any previously shared referents when it comes to decision processes for definite noun phrases. However, since there is no common-ground constraint on definite reference, we made the observation that, in perspective-taking tasks at least, the privately viewable objects are common ground, to some extent. But all that this means is that objects in private view become salient to some degree whilst the director and participant interact in the task. However their identities are not common ground.

One aim for this chapter is to explore this hypothesis in the experiments, taking our cue from research on children's use of cues to interaction to determine what is shared information. One study in particular, reported in Moll et al., (2011) has an experimenter first interact with a two-year old child and two novel objects. In a second phase, the adult either leaves the room while the child is given a third object to play with, replicating the procedure of Tomasello & Haberl (2003), or remains only to have her view of the third object blocked by a barrier. Moll et al. found that when the Experimenter remained in the room, the child was no longer able to infer that the third toy was novel for her at the test phase. It seems that the continued presence of the Experimenter created in the child the impression that the third object was in shared attention as well. This is the case whether the Experimenter continued to communicatively engage with the child about the object from behind a barrier or whether they sat silently behind a barrier, close to the child but without an ability to make eye contact.[23]

The experiments reported in this chapter will test to see whether shared engagement relating to the privately viewable objects in a perspective-taking task makes the task more difficult for participants.

2. Theory of Mind in on-line referential processes

Recall that in our review of child research in Chapter 1, we noted that the result in Moll et al. (2011) answered a question raised by the earlier result reported in Tomasello and Haberl (2003), about whether children succeed in the task by considering the experimenter's individual perspective or their shared perspective. Moll et al.'s result suggest that children at this young age rely more on shared information than information about what a person may know or not know. We assume that through development, people get better at integrating information about other's beliefs and desires, in some cases overriding the salience effect of common ground.

As discussed at length in the last chapter, we see a critical role for the integration of Theory of Mind inferences (about what the speaker is ignorant of) in perspective-taking tasks. We argued that this position differs from the standard Constraint-Based view (see Brown-Schmidt & Hanna 2011), which follows Clark (1996) where conversation is transacted via

[23] See Knoblich et al. (2011) for a review of research on how co-presence itself can lead to a form of contingent reactivity.

what is common ground and what is not. Definite referents are constrained to be in common ground, and new information is constrained to be outside of common ground. At best, information about what an interlocutor may know or not know plays a supportive role in maintaining a representation of ground information.

To date, no study had directly explored the role of Theory of Mind inferences in the perspective-taking tasks. Beginning with the study reported in Keysar et al. (2000) and many subsequent studies from Keysar's lab and others (such as Hanna et al. (2003) and Heller et al. (2008)) perspective taking tasks can be successfully completed by a participant that focusses their attention on what is established as common ground via a co-presence heuristic (Keysar et al., 2000; i.a.) or a previous discourse heuristic (Hanna et al., 2003; i.a.). Although it is open for a participant to infer that if the director does not know the identities of the private objects (e.g. that there is a candle in private view), then they cannot use those identities in a definite description of those objects, it is not necessary for participants to use this kind of ignorance inference to successfully complete the task.

The idea for Experiment 3, reported below, is that if we manipulate the cues to the director's ignorance of the identities of the common ground objects, then we should see an effect of this manipulation in on-line processing, to the extent that participants are able to integrate cues to speaker's ignorance in the perspective-taking task. That is, if participants largely rely on common-ground heuristics to complete the task, there should be no effect of a stronger cue to the speaker's ignorance. However, if participants rely on Theory of Mind inferences, then when there is more information in the interaction about the director's ignorance, we should see better performance.

Thus, in the first experiment reported below, we manipulated cues to the speaker's ignorance of the privately viewable objects. To do this, we adapted the procedure employed in the first two experiments (see Chapter 2). In that procedure, the participant's private objects appear on a side shelf prior to the appearance of the commonly viewable objects. The participant is asked to move the objects to private grid positions of their own choosing and then signal that they are ready to view the common objects. In Experiment 3 below, we included a phase of the procedure between the appearance of the private objects and the appearance of the common objects where the participant is presented with a statement about the superordinate category of the objects (it would be one of either animals, shapes or fruits). In one condition, the director guesses the category and the participant replies whether their statement is True or False. In another condition, a statement about the categories appears on the participant's screen. Again the participant is required to judge the statement True or False. In both cases, the rate of True statements is at chance (1/3). Thus in one condition, the director's behaviour (guessing the superordinate category) highlights their ignorance about what kinds of objects the participant can see. In another condition, although the participant responds to the same statements in the same way, coming from the computer program that controls the experiment, the statements appear more as tests set by the experimenter.

3. Experiment 3

3.1. Method

3.1.1. Participants

40 participants (7 male and 13 female in the guess condition; 9 male and 11 female in the test condition) were recruited from UCL's subject pool. They were blind to purpose of the study. All were paid or given course credit for their participation. One UCL student was recruited to play the speaker role.

3.1.2. Stimuli and design

Participants were given instructions to move objects around a 3*3 grid as in Figure 4.1, where three objects were in common view between the instructing speaker and hearer and three objects were placed in grey positions so that they were only visible to the hearer (privileged view). Except for the training phase where participants were given a turn at being a director, all instructions were given by a confederate speaker.

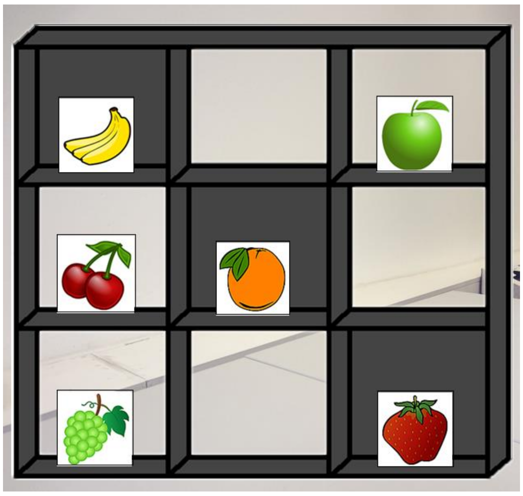


Figure 4. 1. The example display for the main game grids in Experiment 3.

Participants could see any objects in all the nine positions. The instructor could see the grid from the opposite side and the three objects in the private grey grids are not visible (here, the banana, the orange, and the strawberry).

For each trial, a new set of objects appeared on the display. Each such set contained objects belonging to one of three categories: fruits, animals or geometric shapes. There were ten different objects for each category and some were paired in the same subtypes. For example, among the ten fruits, there were two apples, two bananas, two oranges, and for other different fruits. A set of four pictures from different subtypes plus two pictures from the same subtype was randomly chosen for each trial of the experiment, except for the non-competitor condition. In the non-competitor condition all six pictures were from different subtypes. For instance, a non-competitor condition trial might contain a picture of a banana, cherries, grapes, an orange, an apple and a strawberry, while for a matched competitor condition the strawberry was replaced by a paired apple (Figure 4.2). All filler items contained similar objects as the competitor condition where two objects were paired.

A trial consisted of one instruction where the participant was asked to move an object to a location. For example, ‘Move the apple to the bottom middle’. There were 20 critical trials and 20 filler items in each association condition. In 10 critical trials (competitor condition), the privileged objects included one that was the same type as (although not visually identical to) the target in the instruction (an apple). In the other 10 critical trials (non-competitor distractor condition), no privileged object was of the same type as the target object. Figure 4.2 below shows initial set ups for a pair of trials in competitor and non-competitor conditions. For 20 filler items, although privileged objects included one that was the same type as a common objects, neither of the twinned objects were the target in the instruction (an apple) ‘Move the grapes to the middle right’.

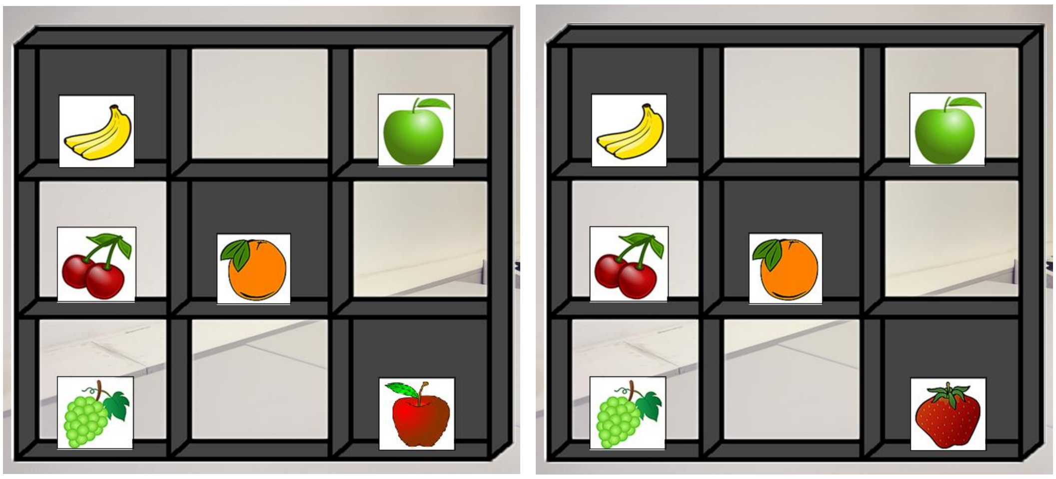
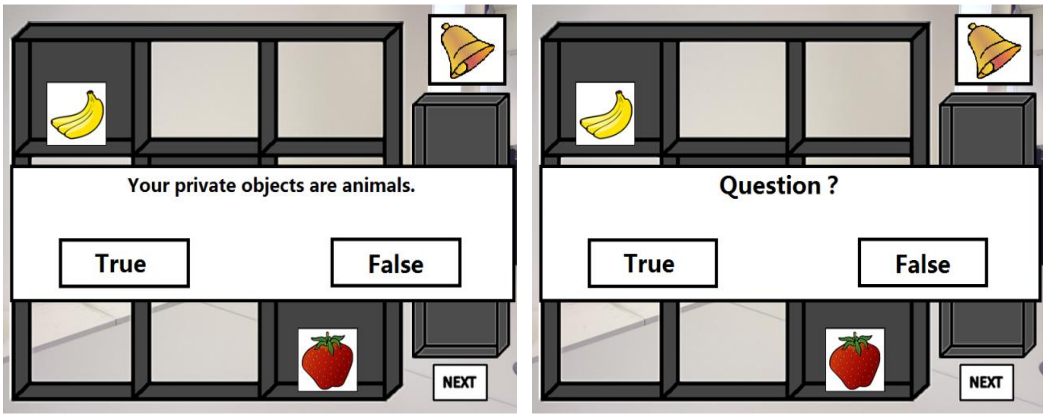


Figure 4. 2. The initial displays for the competitor condition (left) and non-competitor condition (right) in Experiment 3.

In this example, the target green apple had competitor with the same subtype, e.g. a red apple, (competitor condition) or an irrelevant non-competitor, e.g. a strawberry, (non-competitor) in participants' privileged grid.

As in Experiments 1 and 2, each trial begins with the main frame empty and participants' private objects appearing on a shelf at the side. The procedure was for the participant to first move those private objects onto the blocked-off private grid positions on the main frame prior to the appearance of the commonly viewable objects. At this point in the procedure, a between-groups question manipulation was designed. For the Test-question group, participants were asked to make a true/false judgement on a statement provided by the program on the screen (eg. Your private objects are animals. True/False). They were required to click on the true/false button on the screen (Figure 4.3.a). In a Guess condition, the director had been instructed to guess the category of the private objects verbally and participants need to both click on either 'True' or 'False' on the screen (as in the Test question condition) but also give a verbal reply to the speaker including both saying 'True/False' and what categories they were (Figure 4.3.b). The trial order and categories in questions were the same in both conditions, so in the Test-question condition the program pre-set the questions and in the Guess condition the speaker read the questions from a list. One third of the questions required a ‘True’ answer and two thirds, a ‘False’ answer. This ratio corresponded to a chance guessing pattern since there were three categories: fruits, animals, and shapes.



a. Test Condition b. Guess Condition

Figure 4. 3. The displays for question types in Experiment 3.

In the Test condition, participants judged the statement by clicking on the true/false button. In the Guess condition, participants received the guess question from their interlocutor and replied both by clicking on the screen and answering verbally.

To sum up, the experiment had a 2 (competitor) * 2 (question type) design. Altogether, each session with a participant involved 40 trials: 20 critical items and 20 fillers. The order of presentation of trials was randomised. Question type was a between subjects manipulation, so a participant either had only written questions (Test condition) or verbal questions (Guess condition).

3.1.3. Procedure

Eye-tracking Study:

At the beginning of the experiment, a participant was seated in front of a 23" screen fitted with Tobii TX300 eye-tracking equipment in a room. An experimenter brought into the room a confederate who pretended to be another participant. They were told that they would be co-operating a game with one being an instructor and another being an operator, and the instructor should direct the operator to move objects in a shelf presented on each screens. The director sat in front of a 15-inch laptop placed at the left side of the Tobii eye-tracker, so both the participants and the confederate could not see each other's screen, but they can see each other's face and body (see Figure 4.4). The experimenter sat at the corner to introduce and monitor the task.

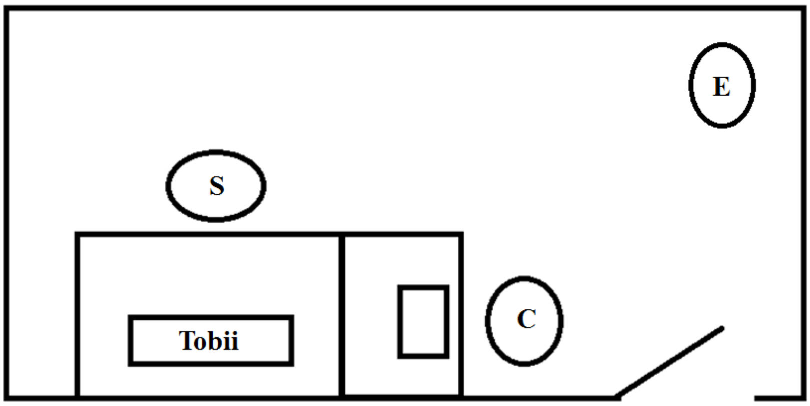
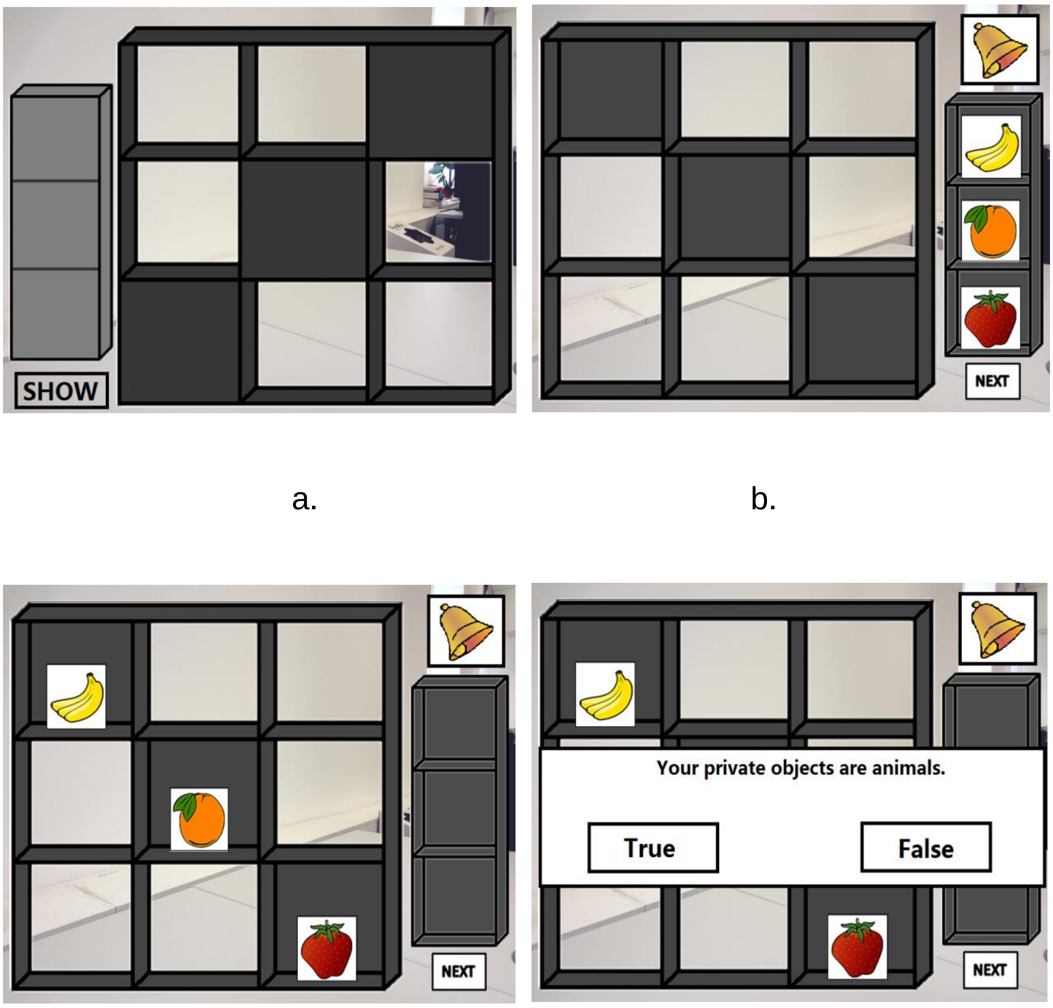


Figure 4. 4. The experimental setup in Experiment 3.

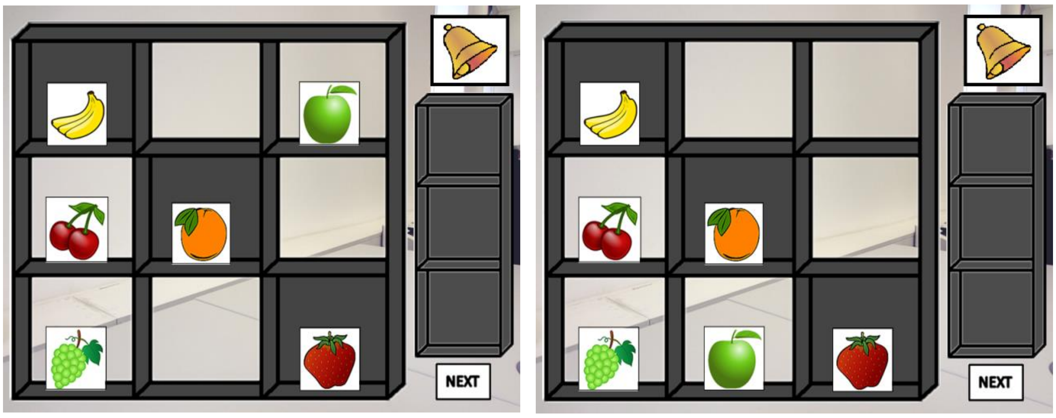
S represents the subject, who was the operator. C represents the confederate, who was the director. E represents the experimenter.

During a training phase, participants were assigned as instructors and the experimenter helped them to give instructions to the operator. The participants were then shown the back of two shelves (see Figure 4.5.a), and the confederate pretended to see the front side. During this phase, participants either showed a programmed Text question or asked a Guess question on the category of operators' private objects, then gave their partner an instruction to move objects according to their star hints. The confederate pretended to move their mouse. After a few trials they switched their roles, and participants were shown the front of the shelves (see Figure 4.5.b).



a. b.

c. d.



e. f.

Figure 4. 5. The procedure for each trial in Experiment 3.

When the participants were the instructors they were shown the back of the shelf (a). When they were the operators they were initially shown the front of the shelf (b), and moved the objects from the small column to grey position of the 3*3 grid (c). The confederate always pretended to see the opposite side of the shelf but in the fact they were watching the same side of the screen. She guessed the category of participants' private objects (d), and then presented three objects in the transparent common positions (e). Finally participants moved an object into a different position according to the confederate's instruction (f).

In Figure 4.5.b, the 3*3 empty grid appeared on the left of the screen. On the right there was a column which contained three objects. In the formal test, the experimenter told participants that the three objects are their private objects that the confederate did not know what they were. The experimenter instructed the participants that they should move these objects to the private grey positions on the 3*3 grid and then clicked on the image of the bell in the top right hand corner of the screen when they are ready (Figure 4.5.c). The confederate then presented the questions (Figure 4.5.d). In the Test-question condition the participants read the question on-screen and just clicked on the answers, without any interaction with the director; in the Guess condition the participants listened to the director’s guess about the category of their private object and answered the question interactively. It needs to be stressed that the participant plays the role of director in the training phase. In that phase, in the Test-condition group, they learn that the director does not see the question appearing on the addressee's screen, nor do they know of the answer. This contrasts with the Guess-condition, where it is the director that makes the statement about the superordinate category, as a guess.

When participants completed this phase, the confederate showed them three objects in common view on the shelf (Figure 4.5.e), then participants rang the bell again to indicate that they have seen the common objects. Speaker then gives one instruction to move an object to a new location (Figure 4.5.f). When this was completed by the participant, they clicked the 'NEXT' button on the bottom right of the screen and the whole procedure began again for the next trial. Note that, on each new trial, the three private grid positions appeared in different, randomly assigned locations around the nine frame grid. After every two trials, participants had an opportunity to break prior to re-calibration of the eye-tracking equipment.

Autism Quotient Questionnaire:

The Autistic-Spectrum Quotient (AQ) developed by Baron-Cohen et. al. (2001) was administered to explore whether the individual autistic traits predicted the task performance. The motivation for using this comes from Nieuwland et al. (2010) that reports an early effect of pragmatic sensitivity, as measured by N400 ERP measure, only for a group with a low AQ score. I include the measure here since, according to my hypothesis, the ability to ignore the occluded distractor in on-line processing would involve access to a Theory-of-Mind inference concerning the instructor's ignorance. It is widely assumed that there are individual differences in abilities to integrate

Theory-of-Mind inferences. To the extent that AQ score reflects Theory of Mind ability, information about this score may enable us to identify at least a sub-group for whom the manipulation here has an effect.

The questionnaire consisted of 50 questions assessing five areas: social skill, attention switching, attention to detail, communication, and imagination. Score 1 point is aggregated if the subject choose an answer with autistic property, so the total score ranges from 0 to 50, and a higher score represented more autistic traits, implying less social ability. The AQ score is not diagnostic but the score 32+ is suggested to be a clinically significant standard of the autistic traits.

Effects of Strategies or Learning:

An aim of this study is to examine the role of two factors in automatic, on-line processes. We saw in Experiments 1 and 2 that participants may adopt strategies or that behaviour can change over time. There were specific factors involved in those experiments that are not involved here, to do with there being instructions to the private objects, encouraging attention to those areas during the instruction. Nevertheless, we included Blocks as a factor in the analysis to check if eye-gaze patterns change over the course of the experiment.

3.2. Analysis and Results

3.2.1. Data Processing

Eye-movements that were initiated during the auditory instruction were processed according to the critical word (‘apple’) onsets for the purpose of aggregating the location and duration of each sample from the eye tracker. For analysis, we removed any sample that was deemed ‘invalid’ due to blinks or head movements. The spatial coordinates of the eye movement samples (in pixels) were then mapped onto the appropriate object regions; if a fixation was located within the square surrounding an object, it was coded as belonging to that object, otherwise, it was coded as background. Target and competitor items were identified on a by-trial basis. In competitor trials, the competitor was the privately viewable object that was of the same type as the target (e.g. the privately viewable apple in the frame on the left side of Figure 4.2). In the non-competitor trials, the target was the object that occupied the same position as the competitor (e.g. the privately viewable strawberry in the frame on the right side of Figure 4.2).

Probabilities of fixating the Target or Competitor/Non-Competitor objects as a function of time were analysed in terms of a target advantage score, (ln(Ptarget/Pcompetitor). Target advantage scores were analysed for three time regions. A preview region began 200ms before, and ended 200ms after, the onset of the critical word (‘apple’). The preview region extends to 200ms after the critical word onset in accordance with standard assumptions about the time to program and launch an eye movement (Hallett, 1986). Following the preview region, our analysis looks at two 300ms regions, 200-500ms and 500-800ms. Following Hanna et al. (2003) the overall 600ms region of interest was chosen to correspond to the average critical word length (between five hundred and six hundred milliseconds). After this period participants' eye movements were often directed toward the mouse curser. The regions were identified and synchronised for each participant on a trial-by-trial basis, relative to the onsets of the critical word.

The average AQ scores in Test condition was 17.7, with a maximum 32 and a minimum 1; the Guess condition average was 18.15, with a maximum 28 and a minimum 7. Those who had a score less than 19 (median of the AQ scores) were allocated into Low AQ group, while others were High AQ group.

3.2.2. Analyses

For each participant (respectively item) and condition, we calculated the average target advantage score over the 17ms time slots per analysis region. Figure 4.6 shows a plot of the time course of target advantage for the two groups. Figure 4.7 shows the average target advantage score for each of the competitor and non-competitor conditions for each of the three windows. And Figure 4.8 shows the proportion of fixations to the target and the competitor in conditions.

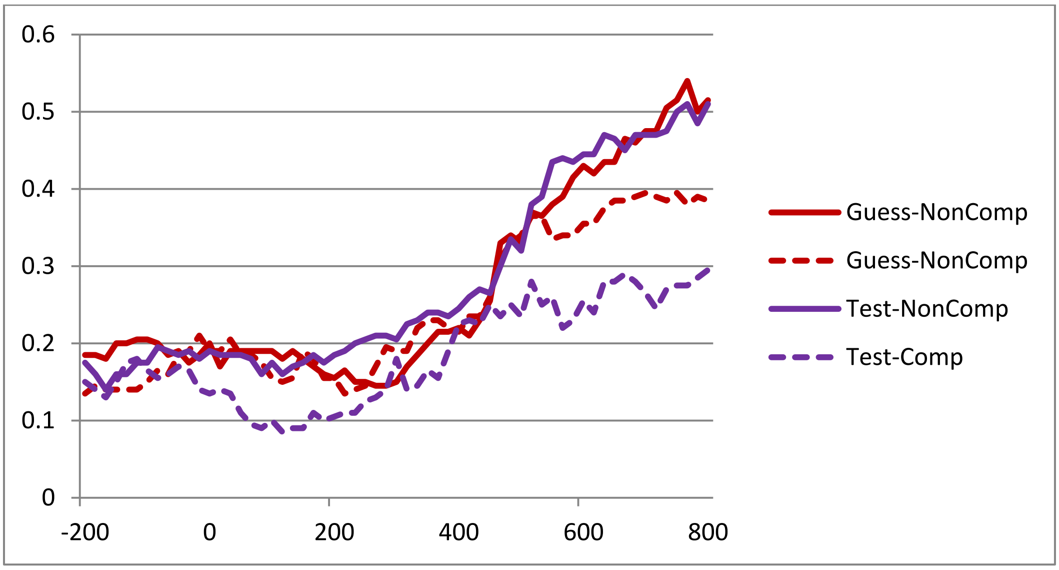
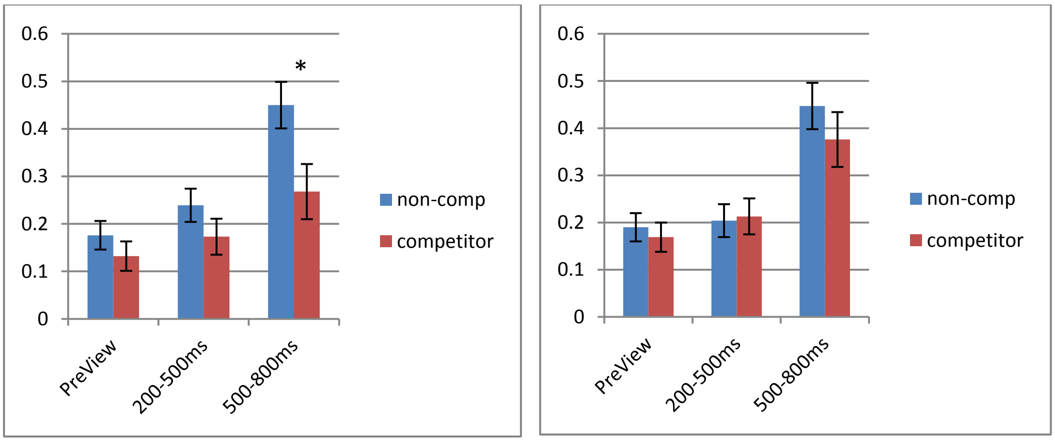


Figure 4. 6. The time course of the average target advantage score in Experiment 3.

The show the TA scores in the guess and test condition from the 200 ms before the onset of the critical word to 800 ms after the onset.

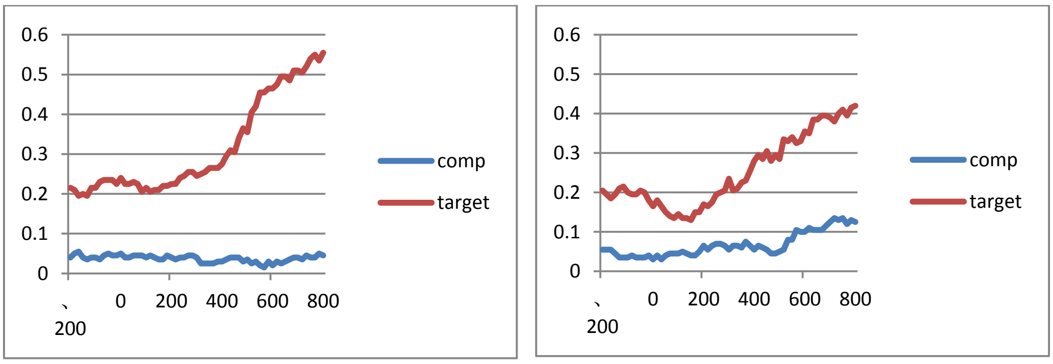


Test condition

Guess condition

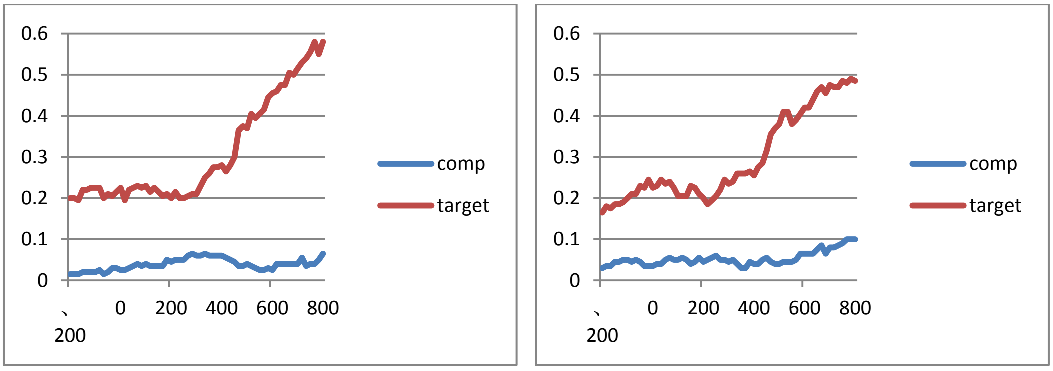
Figure 4. 7. The average target advantage score for the three time windows in Experiment 3.

The charts compared the competitor and non-competitor conditions in the Test (left) and Guess (right) conditions. Error bars indicate standard error of by-participant means.



Test – Noncompetitor

Test – Competitor



Guess – Noncompetitor Guess – Competitor

Figure 4. 8. The proportions of looks the target and the competitor in Experiment 3.

Looking at the graphs in Figures 4.6 and 4.7, we can see an interactive pattern emerge, particularly in the 500-800ms time window.

The results of a 2(competitor vs. non-competitor) * 2(guess vs. test) * 2(high AQ vs. low AQ) * 2(first block vs. second block) mixed ANOVA are as follows.

Preview and 200-500ms window:

In the preview window there were no main effects or interactions among the competitor types, question conditions, AQ scores and blocks (All Fs < 2, all ps > 0.2).

In the 200-500ms window, there was only an effect of AQ due to larger TA in the high AQ than the low AQ group in the by-item analysis, F1(1,72)=2.683, p=0.106, η_p^2=0.036, F2(1,32)=6.488, p<0.05, η_p^2=0.169. No other main effects or interactions were found (All Fs < 3, all ps > 0.1).

This is a perhaps unexpected result since a higher AQ score means lower ‘sociability’. We return to this below.

500-800ms window:

In the second time window (500-800ms), there was significant effect in the competitor types, F1(1,72)=20.872, p<0.001, η_p^2=0.225, F2(1,32)=9.900, p<0.01, η_p^2=0.236, due to the target advantage score being larger in the non-competitor condition than the competitor condition. There was also a significant effect in the AQ scores, F1(1,72)=8.480, p<0.01, η_p^2=0.105, F2(1,32)=21.058, p<0.001, η_p^2=0.397, due to the larger TA scores in the high AQ groups than the low AQ groups.

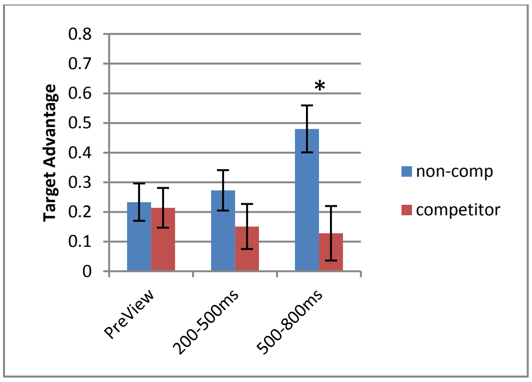
There was an interaction among the competitor types and the question types in the by-subject analysis, F1(1,72)=4.027, p<0.05, η_p^2=0.053, F2(1,36)=2.076, p=0.159, η_p^2=0.061. Simple effect analyses revealed in the test condition participants had larger target advantage scores in the non-competitor condition than in the competitor condition, F1(1,72)=21.617, p<0.001, η_p^2=0.231, F2(1,32)=10.521, p<0.01, η_p^2=0.247, but no significant difference was found in the guess condition, F1(1,72)=3.282, p=0.074, η_p^2=0.044, F2(1,32)=1.454, p=0.237, η_p^2=0.043.

There was also an interaction between AQ scores and the question conditions, F1(1,72)=4.416, p<0.05, η_p^2=0.058, F2(1,32)=12.014, p<0.05, η_p^2=0.273. Simple effect analyses showed that the low AQ group had larger TA scores in the Guess condition than the Test condition, F1(1,72)=5.442, p<0.05, η_p^2=0.070, F2(1,32)=12.709, p=0.001, η_p^2=0.284, but the high AQ group have same TA scores in both question conditions, F1(1,72)=0.408, p=0.525, η_p^2=0.006, F2(1,32)=1.787, p=0.191, η_p^2=0.053.

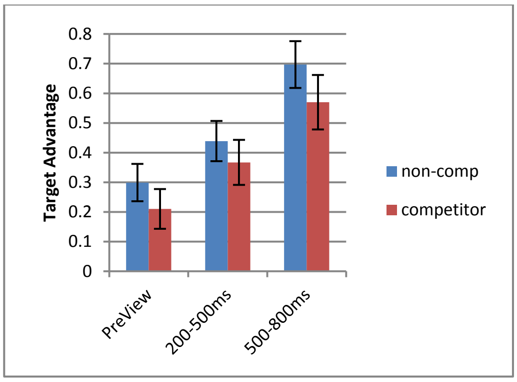
To sum up these results, the first thing to note is that there were no effects by Block at all. This contrasts with the first two experiments but was expected since the procedure here was closer to the normal Perspective-taking procedure and was not designed to induce memory-based or strategic effects.

Next, in the 500-800ms window, the Competitor-by-Question type interaction revealed the predicted positive effect of making the speaker's ignorance salient: Target Advantage score in the Guess condition did not differ between the Competitor and Non-Competitor conditions, whilst the there was a difference in the Test condition, with Competitor trials giving rise to lower Target Advantage.

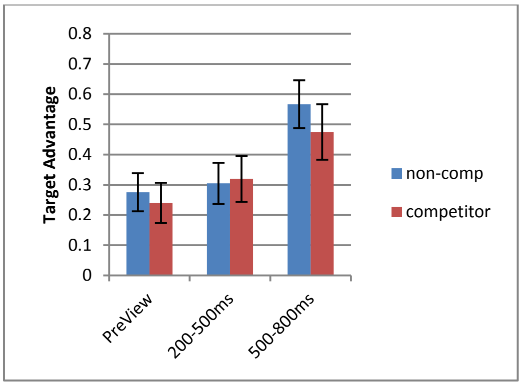
In line with the surprising result in the 200-500ms time window we again find an unexpected main effect of AQ score in the 500-800ms time window. In addition, we find an AQ-by-Question type interaction, showing that the Low AQ group had worse performance on the Test-question condition. To make sense of these surprising results, it is worthwhile considering the graphs in Figure 4.9 below which show the average TA per time window broken for each of the Low- and High-AQ groups.

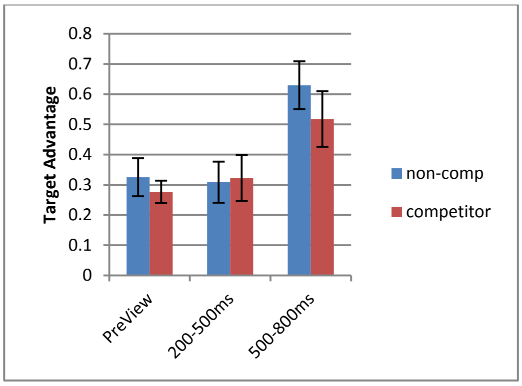


Test - Low AQ



Test - High AQ





Guess - Low AQ Guess - High AQ

Figure 4. 9. The average TA scores for two AQ groups in Experiment 3.

The charts compared the competitor and non-competitor conditions in the Test (upper) and Guess (lower) conditions, and Low AQ (left) and High AQ (right) group. Error bars indicated standard error of by-participant means.

Inspection of the graphs in these figures suggests that the more 'socially able' Low AQ group has more difficulty in the Test condition than the High AQ group.

One hypothesis for this somewhat surprising result lies in the nature of the AQ test itself. Rather than being a direct test of Theory-of-Mind abilities or relevant Executive-Function abilities like Inhibition control, the AQ survey looks at a more rounded picture of the sociability that is appropriate for assessing for Autistic Spectrum tendencies. It is widely known that Autistic Spectrum individuals are able to pass False-Belief tasks, standardly administered to test Theory of Mind (Happe, 1994) but it now becoming clear that implicit social-cognitive abilities differ from neurotypical adults (Schneider et al., 2013). If our procedure in the Guess question condition (making ignorance salient via a standard marker of ignorance) activates a more explicit Theory of Mind ability, this could explain why both groups show comparable Target Advantage in Competitor and Non-Competitor conditions in that experiment. But this does not explain why socially more able individuals find the baseline Test procedure more difficult than the socially less able group. This thesis' conjecture about the perspective-taking procedure, to be developed in more detail below, is that even though the specific identities of the private objects are unknown to the speaker, their presence is shared information, in shared attention. As discussed throughout the thesis, processing biases makes objects that are, or have been, in shared attention more salient or available for referential processes. But this is likely to be a graded phenomenon and it could be that individuals differ in the extent to which they process information as shared. In fact, it could be part of the repertoire of more implicit social cognitive abilities to process information as shared, using low-level social cues. Thus, one explanation for why the Low AQ group had more difficulty on the Test condition is that the private-view objects are more 'shared' for them and hence more salient as potential referents in the instructions. We return to this discussion after Experiment 4.

3.3. Discussion

We adapted the standard perspective-taking procedure to include a phase where the director and participant could interact and share attention to the private objects, or not. This happened in the guess question, where the director's guess about the superordinate category of the objects makes salient the director's ignorance of what type of object they are. In the non-interactive condition, the participant simply had to respond True/False to a statement on screen.

We analysed eye-gaze data within a 600ms time window. As the mean length of the critical noun is about 500ms, this window is little more than the time it takes to process the noun. We found that within this relatively short time window, participants showed a differential effect of condition. In the Guess condition participants were better able to ignore the distractor object in private view, to the extent that there was no difference in bias formation compared to when only irrelevant distractors were in private view. In the Test condition, by contrast, participants' gaze data revealed a competitor distraction. Compared to the non-competitor distractor condition, participants had achieved significantly less Target Advantage in the Test condition.

It is important to note that the interactive, guess condition did not involve making the identities of the private objects common ground (i.e. whether they were apples or grapes etc.). So, according to the common ground constraint proposed by Clark & Marshall and widely adopted in the literature, these objects are not common ground for definite descriptions used in the experiment ('the apple'), since information about these identities is not common ground. Our conjecture is that there is no common ground constraint on definite reference that requires the descriptions of the objects already be common ground, but that common ground is a result of cognitive biases triggered by social-cognitive cues. Common ground information based on prior interaction becomes salient for interlocutors as a result of low-level memory mechanisms (Horton & Slaten, 2012) and is a graded phenomenon (Brown-Schmidt, 2012).

In contrast to Constraint-Based models, we see a central role for Theory of Mind inferences in perspective taking tasks. That the speaker is ignorant of the identities of the objects in private view, although their existence is, to an extent, common ground, is a fact that needs to be integrated into referential decision processes. By contrast, a proper implementation of Clark's framework (see Clark, 1996), would see inferences about ignorance of objects being included into decisions about common ground. I.e. one would infer that what is not known by an interlocutor could not be common ground.

In this experiment, we have illustrated the rapidity with which information about the speaker's ignorance can be integrated into decisions about reference. In the Guess condition, the director's guessing behaviour is a strong cue to their ignorance of the private object's identities. Compared to the baseline Test condition, we see the effect of this cue in a time window beginning a mere 300ms after the onset of the critical word.[24] To compare this timing it is instructive that in the study reported in Keysar et al (2000) the average first looks to the target occur at 1463ms. In a study employing similar materials, a similar sized grid but a simpler task, Hanna et al. (2003) report a significant difference between target and competitor bias from 400ms; but a continued interference from the competitor compared to the non-competitor distractor across the same 200-800ms time frame we examined here. These latter results are comparable to our Test condition results. Thus, although it is difficult to compare timings across different experiments, it seems that highlighting speaker's ignorance of private objects clearly improves performance in a time window that is inconsistent with the hypothesis that so-called higher-order Theory of Mind information is costly and/or time-consuming to integrate. Experiment 3 thus provides further evidence against the perspective-adjustment model, which assumes that information about the speaker's ignorance is costly to compute and integrate into referential decisions. In fact, Experiment 3 demonstrates precisely that the integration of Theory of Mind inferences is a graded phenomenon, being something that is supported by contextual information, as proposed in Ferguson & Breheny (2011).

In sum, Experiment 3 shows that cues to speaker ignorance leads to improved performance on a perspective-taking task. This is the first experiment to directly test the effect of such cues on performance. The results thus indicate an important role for Theory of Mind inferences in such tasks. The results are also counter-indicative of proposals set out by

[24] Recall that it takes 200ms to launch an eye movement having processed some linguistic stimulus (Hallet, 1986), so gaze data recorded at 500ms after the absolute onset reflects processes 300ms after the onset of the word.

Constraint-Based theorists that tracking cues to common ground is sufficient for good performance on this task.

An alternative account of the effect of guessing in this experiment would turn on a second difference between the Test and Guess conditions. Recall that in the Test condition, the participants simply respond to a statement that appears on their screen. They are aware the Director does not see this statement or their response. By contrast, in the Guess condition, there is a full interaction about the superordinate category of the private objects. From the Constraint-Based perspective, it could be argued that the simple fact of greater interaction between the participant and director on each trial leads to better performance. In the only other study of the effect of interaction on performance, Brown-Schmidt (2009a) reports that participants were more responsive to ground information when they interacted with a live director compared to when they were instructed via an audio recording. It is important to consider Brown-Schmidt's studies in detail to see that in fact the results are not only consistent with the results reported here but with the overall thesis that ground information and ToM information are separate.

Brown-Schmidt (2009a) reports on studies that explore the phenomenon of so-called entrainment or conceptual pacts (Brennan & Clark, 1996; Garrod and Anderson, 1987). The paradigm exploits the fact that in conversation, people tend to re-use descriptions for an object, rather than change. Thus, if one speaker calls an object a metal pipe, both interlocutors will tend to continue to use that description, other things equal. In studies of the type that Brown-Schmidt reports, one tests the effect of changing the description. E.g. from, 'metal pipe' to 'silver tube' and/or changing the speaker. In a visual-world study using a live confederate, Brown-Schmidt finds that when the speaker changes but the description does not, bias to the target does not form as quickly as when the original speaker re-uses the entrained upon description. But when the original speaker changes the description, target bias is slower to form than when a new speaker uses the same novel description. These effects were found in an early time window, 200-600ms after the onset of the critical linguistic input. Thus Brown-Schmidt's results disconfirm the claim in Kronmüller & Barr (2007) that partner-specific effects should take time to emerge, whilst ego-centrically derived precedence effects should be evident at the earlier time windows. When Brown-Schmidt conducted the same study using audio recordings (as Kronmuller & Barr had done) instead of live directors , the effects disappear at the early time windows and only re-emerge at later times, comparable to those reported in Kronmuller & Barr (2007). Brown-Schmidt's conclusion from these results is that greater live interactivity between speakers provides a greater resource for establishing ground information, leading to the early effects of discrimination among speakers.

Could it then be the case that greater interactivity in Experiment 3 above was the cause of the better performance? Interactivity is not something that is controlled for in Experiment 3, over and above the design providing evidence for the director's ignorance of the private objects. Thus, we require a follow-up study that can serve as a control for this factor. This will be presented as Experiment 4 below.

Before we come to the details of Experiment 4, it is worth saying a few words about the studies in Brown-Schmidt (2009a) and how their results could be accounted for in the framework advocated in this thesis. The first thing to note about entrainment research is that it examines a phenomenon that can be accounted for simply in terms of a common ground mechanism of the kind proposed in this thesis. The first effect reported, where maintained description/maintained speaker leads to greater early bias than maintained description/different speaker, can be explained simply in terms of cued memory mechanisms. The compound cue of speaker and description provides a stronger activation of the target than the simple cue of description. The second effect can be explained in terms of common ground plus a phenomenon discussed at length in

Pickering and Garrod (2013) whereby production can influence comprehension. In the entrainment task, participants would be liable to pre-label objects according to previously used descriptions, using the speaker of those descriptions as a cue. Thus, the ‘pre-emption’ effect seen in early time windows in Brown-Schmidt’s study could be the product of a low-level common ground mechanism and the fact that production processes feed into comprehension.

While greater interaction can improve performance on entrainment tasks, our discussion of the role of common ground in the last chapter led us to propose that interaction, per se, might lead to worse performance on the perspective-taking task. Our proposals about how social-cognitive mechanisms operate separate out mechanisms for Theory of Mind inferences and mechanisms for maintaining common ground, based on cues to shared attention or shared experience. In Experiment 3, the kind of interaction we introduced highlighted the director’s ignorance of the types of objects in private view for the participant. However, we conjectured in Chapter 3 that the general procedure for the perspective-taking task makes the objects in private view common ground, to some extent. Thus, an interaction between director and participant that increased the amount of shared attention to these private objects but that did not highlight the director’s ignorance of their identities, would, according to the proposals in this thesis, increase competitor interference, relative to a comparable non-interactive condition.

4. Experiment 4

In the Experiment 4, we intend to explore whether only social engagement affects people’s perspective taking. In the Experiment 4 we adapt the procedure from Experiment 3 to test this hypothesis, and manipulate the social engagement in three dimensions: Interactive, Low-interactive, and

Non-interactive. In the pre-test phase of the guess trial in Experiment 3, the director makes a guess about the superordinate category of the private objects (making their ignorance salient). In an Interactive condition in Experiment 4, the director is allowed to select the superordinate category of the objects and announces their selection. In a Low-interactive condition, there is no verbal communication between interlocutors but participant and director are co-present and each given a message about the superordinate category of the private objects at this preliminary phase. This condition is comparable with the Test condition in the Experiment 3 due to a similar degree of the social engagement. Physical co-presence can convey some social cues. Simple co-presence has been shown to induce shared representations and forms of contingent behaviour (see Knoblich et al. 2011 for a review). Hence a third non-interactive condition was established with no verbal conversation at this preliminary phase, no shared message and no co-presence.

It has been shown that the social engagement facilitates young children to recognise other's knowledge (Moll & Tomasello, 2007), but the co-presence of conversational partners can also lead to an overestimation on their mutual knowledge, because the privately viewable objects are in common ground to some extent (Moll et al., 2011). So the results of the present experiment will be considered as a control study of Experiment 3. Our prediction is made based on the graded dimensions: participants may resort more to their privileged knowledge when their partner is present than absent, and when they have a conversation with their partner than without communication.

4.1. Method

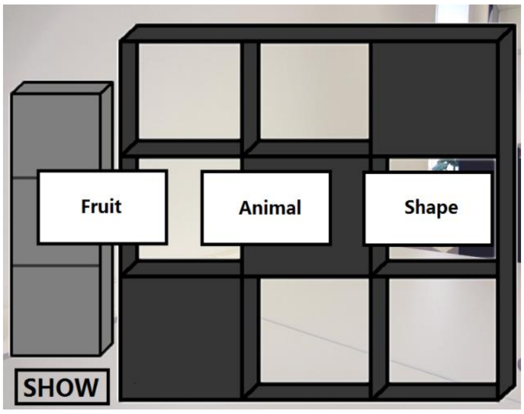
4.1.1. Participants

60 participants (12 male and 8 female in the interaction condition; 9 male and 11 female in the low-interaction condition; 11 male and 9 female in the low-interaction condition) were recruited from UCL's participant pool. They were blind to the purpose of the study. All were paid or given course credit for their participation. One UCL student was recruited to play a role of a confederate.

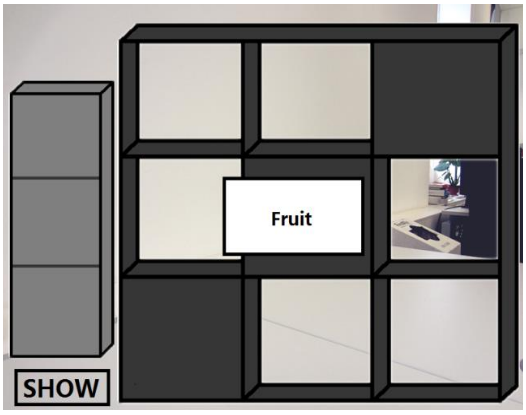
4.1.2. Stimuli and design

Experiment 4 presented the same shelves and object pictures as Experiment 3. In the present one, instead of guessing the object category, the category was assigned by either the instructor or the computer. Participants heard or read a statement before their private objects were presented. The type of interaction was operationalized in a between group manipulation. In an Interactive condition, the confederate director verbally assigned the categories, for example, 'I'll show you fruits, ok?' Participants were required to acknowledge their agreement. Then the director clicked on one of the choice shown in Figure 4.10.a. In a Low-interactive condition, a notice of the category appeared on both sides' screens. Thus both the director and participant could to see the notice, for example, "fruit" shown in Figure 4.10.b, but there was no verbal interaction about this. In an Non-interactive condition, the director sat in a different room, and no notice appeared on her screen. In this way there was neither interaction nor even a shared message about the participant's privileged objects prior to the test phase. The trial order and categories were the same for all conditions.

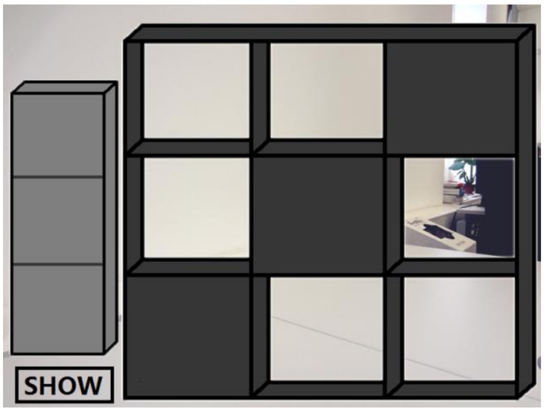
There were 10 competitor trials, 10 non-competitor trials and 20 filler trials in each condition. Competitor trials had three privileged objects which included one that was the same subtype (although not visually identical to) as a common object, and the common object would be a target object. Non-competitor trials had six objects in different subtypes. Filler items also had twinned objects as competitor items, but neither of them was the target.



a.



b.



c.

Figure 4. 10. The initial displays in the instructor's screen in Experiment 4.

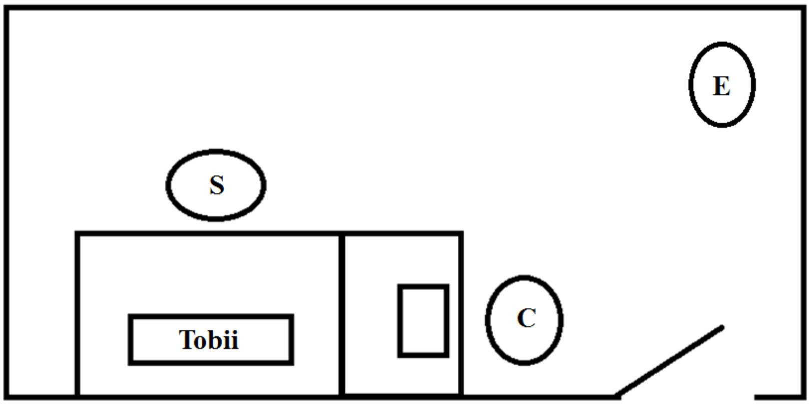
Participants were shown the above display in practice trials where they played the role of instructors. The upper left figure (a.) represented the Interactive condition, where the instructor needed to read and click the notices to demonstrate objects in the participant’s shelf. The upper right figure (b.) represented the Low-interactive condition, where a statement about the category of objects appears on both the instructor’s and participant’s screen. The lower figure (c.) represented the Non-interactive condition, where the director had no information about the objects.

To sum up, the experiment had a 2 (competitor) * 3 (interactivity) design. Altogether, each session with a participant involved 40 trials: 20 critical trials and 20 fillers. The order of presentation of trials was randomised. Interactivity was a between subjects manipulation. So a participant was given prior notification of the category of objects (animal, shape, fruit) either via verbal notices (Interactive) from a co-present director or via a shared written notice while the director was co-present (Low-interactive condition) or via a private (unshared) notice and the director was stationed in a separate room (Non-interactive condition). This announcement coincided with the presentation of the participant’s private objects and thus prior to the trial task.

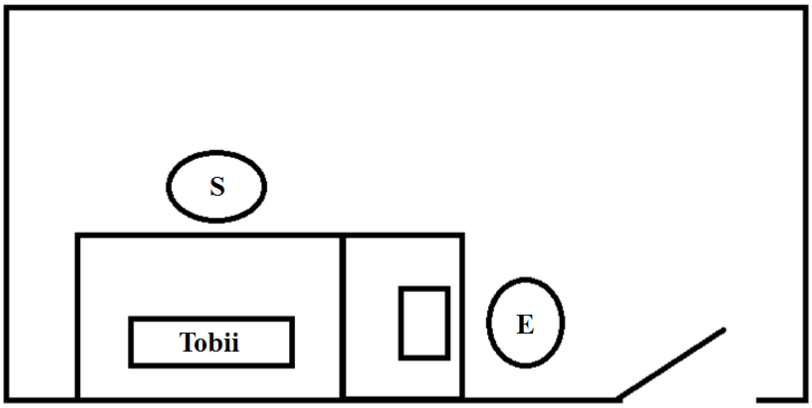
4.1.3. Procedure

The present experiment was performed in the same lab as the first experiment. The confederate pretended to be a participant. At the beginning of the experiment, a participant was seated in front of a 23” screen fitted with Tobii TX300 eye-tracking equipment in an eye-tracking lab. An experimenter brought into the room a confederate who pretended to be another participant. They were told that they would be co-operating a game with one being an instructor and another being an operator, and the instructor should direct the operator to move objects in a shelf presented on each screen.

In the Interactive and Low-interactive condition, the confederate sat in front of a 15-inch laptop placing at the left side of the Tobii eye-tracker, so both the participants and the confederate could not see each other's screen, but they could see each other's face and body (see Figure 4.11). The experimenter sat in the corner to introduce and monitor the task. In the Non-interactive condition, after introducing each other, the confederate was brought out of the room to perform the task in another room. During the Non-interactive sessions, the experimenter sat in front of the laptop as the confederate in the other two conditions.



a. Interactive and Low-interactive conditions

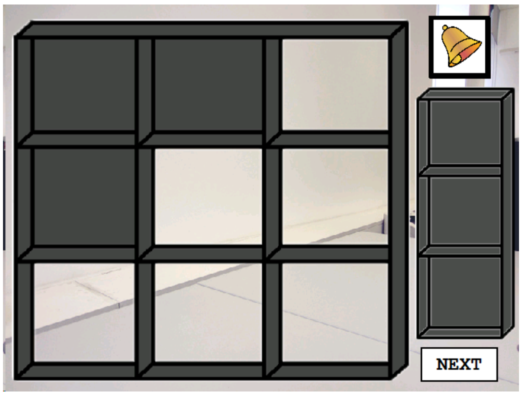


b. Non-interactive condition

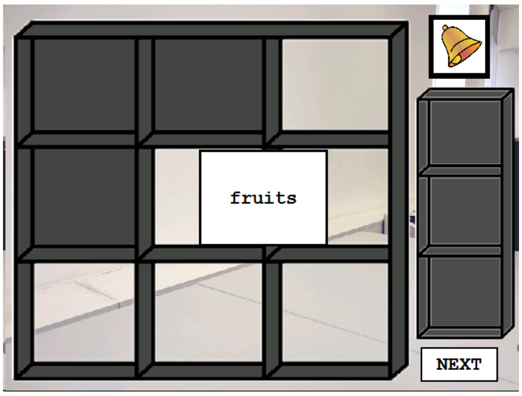
Figure 4. 11. The experimental setups in the three conditions in Experiment 4.

S represents the subjects, C represents the confederate, and E represents the experimenter. The upper display demonstrates the setup of the interaction and non-interaction conditions, and the lower display demonstrates absent condition.

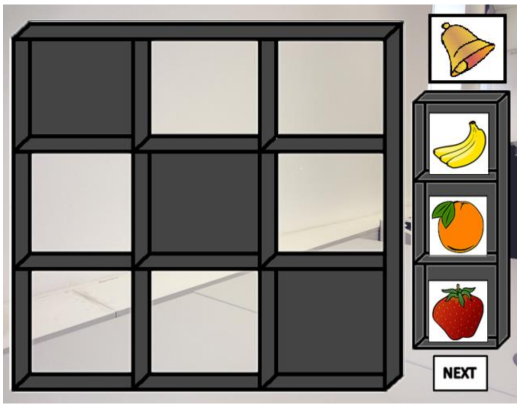
Once testing began, after the training phase, the confederate was an instructor and directed the participant to move objects around a shelf. But they both had opportunity to experience the other's role during a training phase. Similar to previous experiments, in the training phase participants were give three trials as director. They first saw the back of two shelves (Figure 4.12). Then they chose (Figure 4.12.a), or be assigned with (Figure 4.12.b), or no idea about the category of objects (Figure 4.12.c) on their partner's screen, depending on the conditions. Then, participants gave instructions to move objects around the grid, playing the Director role for the training phase.

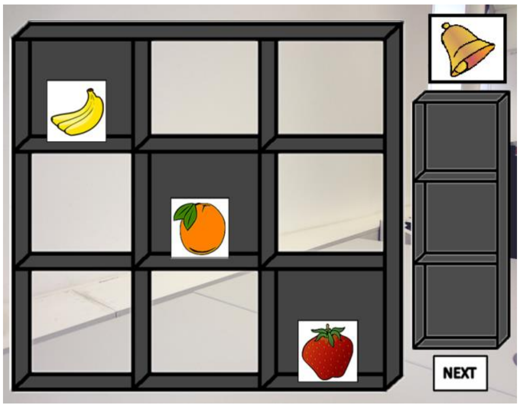


a.

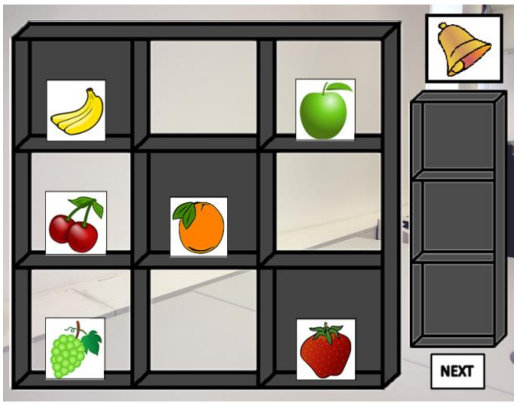


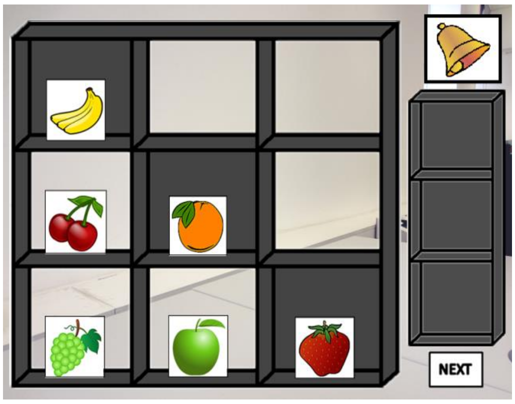
b.





c. d.





e. f.

Figure 4. 12. The procedure for each trial in Experiment 4.

The participant first sees both shelves empty (a), then a category label appears (b), having been announced or not by the Director. This is followed by the appearance of private objects on the smaller shelf to the right of the main shelf (c). The participant then moves their private objects to blocked-off grid positons on the main shelf (d) and rings the bell. At this point, the commonly viewable objects appear (e) and when the participant signals that they are ready, they are given one instruction to move a commonly viewable object by the director (f).

Once roles were swapped, the confederate took over as director and the participant took on the role of operator. In the Interactive condition the director announced to the participants, "this time I'll make it fruit. Ok?", clicking on the label on their screen and then the participants had to verbally acknowledge the decision. Both the Low-interactive and Non-interactive condition lacked communication during this stage. However, in the Low-interactive condition, the participant is aware that the director sees the category of the private objects just as the three private objects appear in the side shelf, according to the pre-announced category. They moved them to the grey privileged positions on the 3*3 grid and then clicked on the image of the bell when they were ready (Figure 4.12.d). When the participant had completed this phase, the confederate presented them three objects in the common view on the shelf (Figure 4.12.e). The participant then clicked on a bell to acknowledge that they saw the common objects. At this point the director gave an instruction to move an object to a new location (Figure 4.12.f). The object was either the target in critical trials or an irrelevant object in fillers. When this is completed by the participant, they click the ‘NEXT’ button on the bottom right of the screen and the whole procedure begins again for the next trial.

4.2. Analysis and Results

4.2.1. ANOVA analyses

Eye-movements were initiated and analyzed in the same way as Experiment 3. The same acoustic onsets and spatial coordinates were coded. Target and competitor items were identified on a by-trial basis. A target advantage score was calculated as dependent variables. Figure 4.13 shows Target Advantage scores over time.

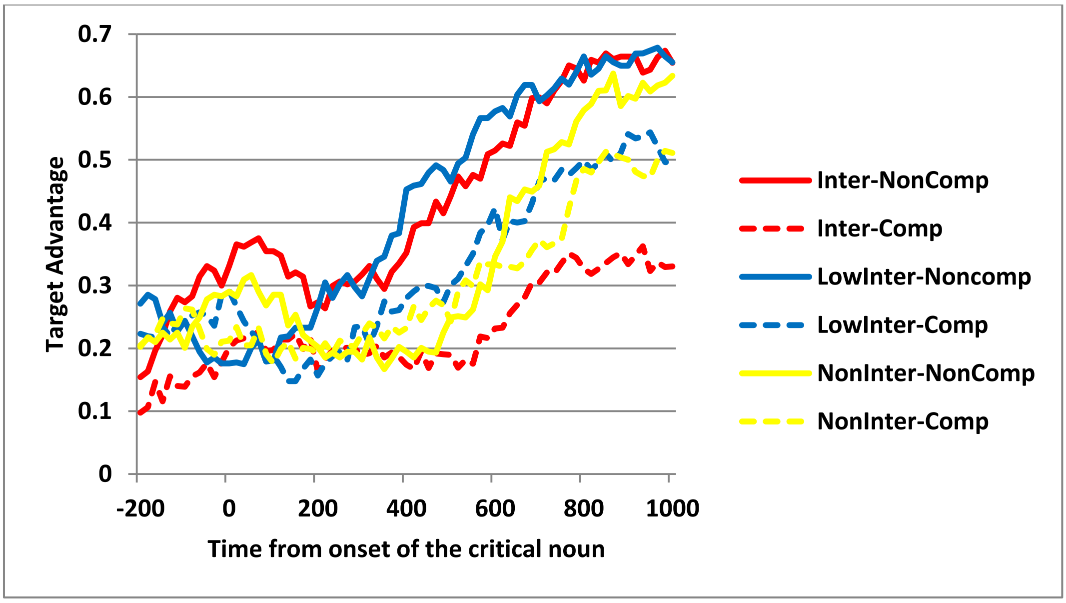
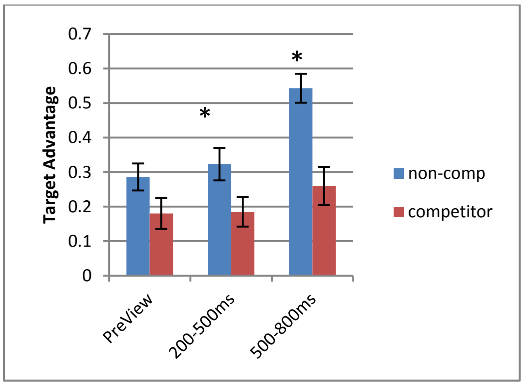
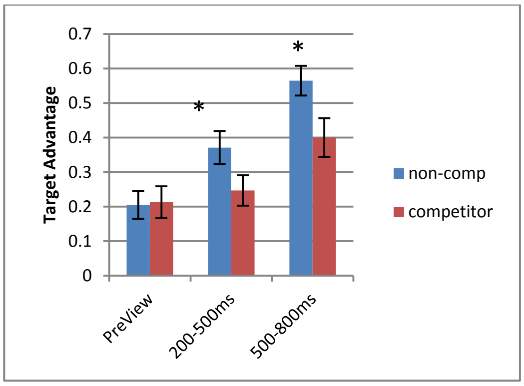


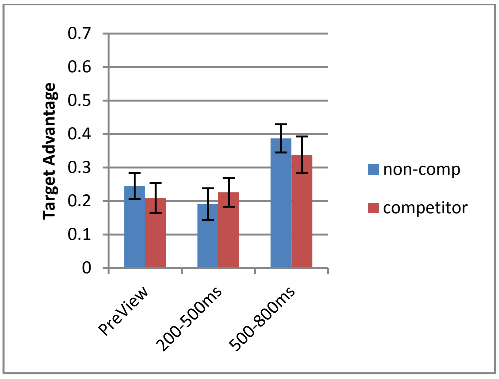
Figure 4. 13. The time course of the target advantage scores in Experiment 4. The solid line indicates the non-competitor items and the dashed line indicated the competitor condition. The same colour of the line represented the same condition.

As with previous experiments, we considered gaze data across three regions: a -200ms-200ms preview region, and two 300ms windows (200-500ms, 500-800ms). Figure 4.14 shows average Target Advantage in the three regions for the six conditions. Figure 4.15 shows the proportion of fixations to the target and the competitor in conditions.





Interactive condition Low-Interactive condition



Non-interactive condition

Figure 4. 14. The average target advantage scores for the three time windows in Experiment 4.

The charts compared the competitor and non-competitor conditions in the Interactive condition (upper left) and the Low-interactive condition (upper right) and Non-interactive condition (lower). Error bars indicate standard error of by-participant means.

The results of a 2(competitor vs. non-competitor) * 3(interactive vs. low-interactive vs. non-interactive) * 2(high AQ vs. low AQ) * 2(first block vs. second block) mixed ANOVA are as follows.

Preview window:

In the preview window, there were no main effects, but there was an interaction between competitor types and blocks, F1(1,108)=5.004, p<0.05, η_p^2=0.044, F2(1,48)=4.950, p<0.05, η_p^2=0.093, and an interaction among competitor, AQ and blocks, F1(1,108)=7.419, p<0.01, η_p^2 =0.064, F2(1,48)=8.421, p<0.01, η_p^2=0.149. Simple effect analyses showed that in the first block, there was simple interaction between competitor types and AQ, F1(1,54)=4.996, p<0.05, η_p^2 =0.085, F2(1,24)=5.000, p<0.05, η_p^2=0.172. Simple effect showed that the low AQ group had the same TA scores in both competitor and non-competitor conditions, but the high AQ group had larger TA score in the non-competitor than the competitor condition, F1(1,54)=11.112, $p<0.01$, η_p^2=0.171, F2(1,24)=11.526, $p<0.01$, η_p^2=0.324. In the second block there was no simple main effect and interactions.

200-500ms window:

In the first time window (200-500ms), there was a significant effect of the competitor type, F1(1,108)=6.173, $p<0.05$, η_p^2=0.054, F2(1,48)=9.180, $p<0.01$, η_p^2=0.161, due to the target advantage score being larger in the non-competitor condition than the competitor condition. There was significant interaction between the competitor type and interactivity, F1(2,108)=3.617, $p<0.05$, η_p^2=0.063, F2(2,48)=5.451, $p<0.01$, η_p^2=0.185. Simple effect analysis showed that in both the Interactive and Low-interactive condition, there were significant simple effects of the competitor types. In the Interactive condition, there was a larger TA score in the non-competitor condition than the competitor condition, F1(1,108)=7.372, $p<0.01$, η_p^2=0.064, F2(1,48)=10.777, $p<0.01$, η_p^2=0.183. In the Low-interactive condition, there was also a larger TA score in the non-competitor condition than the competitor condition, F1(1,108)=5.372, $p<0.05$, η_p^2 =0.047, F2(1,48)=8.425, $p<0.01$, η_p^2 =0.149. But in the Non-interactive condition, no simple effect of the competitor types was found, F1(1,108)=0.559, $p=0.456$, η_p^2=0.005, F2(1,48)=0.879, $p=0.353$, η_p^2=0.018. There was also an interaction between competitor types and blocks, F1(1,108)=6.764, $p<0.05$, η_p^2=0.059, F2(1,48)=10.370, $p<0.01$, η_p^2=0.178. Simple effect showed that in the first block, participant had larger TA scores in the non-competitor condition than the competitor condition, F1(1,108)=12.929, $p<0.001$, η_p^2 =0.107, F2(1,48)=19.532, $p<0.001$, η_p^2=0.289, but in the second block, there was no difference between competitor types.

500-800ms window:

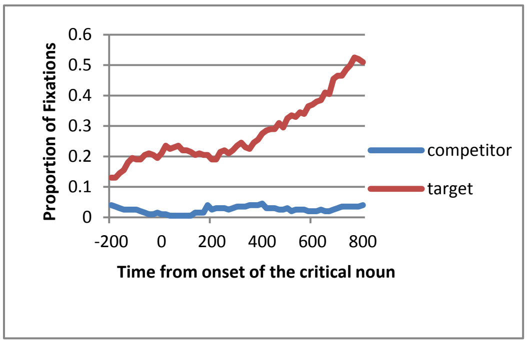
In the second time window (500-800ms), there was a significant effect of the competitor types, $F1(1,108)=23.345$, $p<0.001$, $\eta_p^2=0.178$, $F2(1,48)=41.402$, $p<0.001$, $\eta_p^2=0.463$, due to the target advantage score being larger in the non-competitor condition than the competitor condition The effect of interactivity was also significant, $F1(2,108)=3.240$, $p<0.05$, $\eta_p^2=0.057$, $F2(2,48)=3.979$, $p<0.05$, $\eta_p^2=0.142$. Post hoc analysis showed that compared to the Non-interactive condition, there was a larger TA score in the Low-interactive condition, $F1(1,78)=6.529$, $p<0.05$, $\eta_p^2=0.077$, $F2(1,38)=6.210$, $p<0.05$, $\eta_p^2=0.140$. No difference was found between the High-Interactive and Low-interactive condition, $F1(1,78)=3.016$, $p=0.086$, $\eta_p^2=0.037$, $F2(1,38)=3.276$, $p=0.078$, $\eta_p^2=0.079$, and between the Interactive and Non-interactive condition, $F1(1,78)=0.792$, $p=0.376$, $\eta_p^2=0.010$, $F2(1,38)=0.606$, $p=0.441$, $\eta_p^2=0.016$. It indicated that compared to the Non-interactive condition, participants found the target more quickly and looked longer generally in the Low-interactive condition.

There was also an interaction between competitor types and blocks, $F1(1,108)=5.234$, $p<0.05$, $\eta_p^2=0.046$, $F2(1,48)=9.606$, $p<0.01$, $\eta_p^2=0.167$. Simple effect showed that in the first block, participant had larger TA scores in the non-competitor condition than the competitor condition, $F1(1,108)=25.344$, $p<0.001$, $\eta_p^2=0.190$, $F2(1,48)=45.447$, $p<0.001$, $\eta_p^2=0.486$, but in the second block, there was no difference between competitor types.

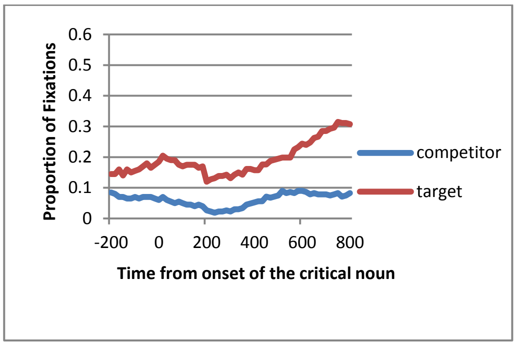
There was interaction between the competitor type and interactivity, $F1(2,108)=4.209$, $p<0.05$, $\eta_p^2=0.072$, $F2(2,48)=41.402$, $p<0.001$, $\eta_p^2=0.463$. Simple effect analysis showed that in the Interactive and Low-interactive conditions, there's a significant effect of the competitor types. Specifically, there was a larger TA score in the non-competitor condition than the competitor condition in the Interactive condition, $F1(1,108)=23.782$, $p<0.001$, $\eta_p^2=0.180$, $F2(1,48)=39.021$, $p<0.001$, $\eta_p^2=0.448$**,** and in the

Low-interactive condition, F1(1,108)=7.398, p<0.01, η_p^2 =0.064, F2(1,48)=14.004, p<0.001, η_p^2=0.226. But no simple effect of competitor types in the Non-interactive condition was found, F1(1,108)=0.599, p=0.441, η_p^2=0.006, F2(1,48)=1.336, p=0.253, η_p^2=0.027.

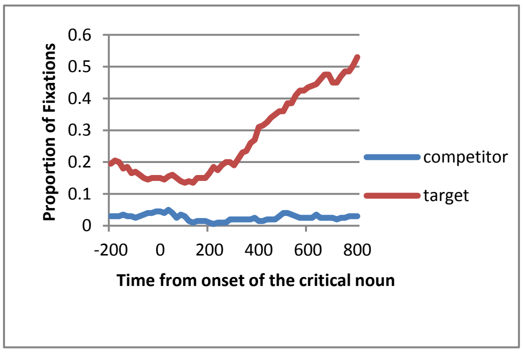
As we will discuss below, these results tend to support the prediction that lesser interactivity concerning the private objects would yield less competitor interference. However, it is possible that the lack of difference between competitor and non-competitor conditions might have arisen due to a much smaller overall bias to the target in the Non-interactive condition – consistent with the main effect of interactivity in the second time window. Figure 4.15 below shows a plot of proportions of looks separately for the target and competitor/non-competitor distractor in each of the three interactivity conditions. Visual inspection of these graphs suggests that the bias to target did not differ drastically between the Non-interactive and the two present conditions. Nevertheless we explored this possibility by a series of t-tests.



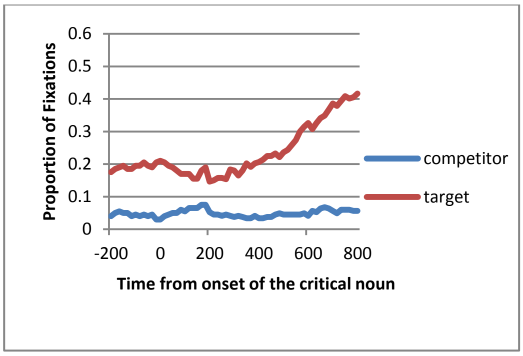
Interactive – Non-competitor



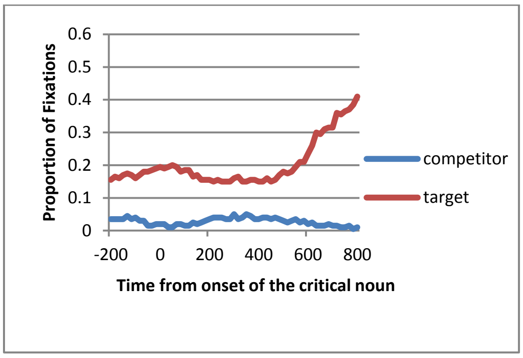
Interactive – Competitor



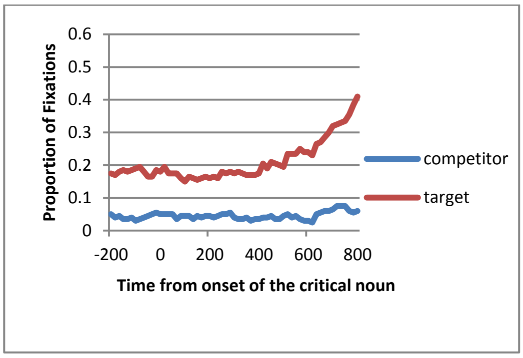
Low-interactive – Non-competitor



Low-interactive – Competitor



Non-interactive– Non-competitor



Non-interactive– Competitor

Figure 4. 15. The proportion to fixation to the target and the competitor in Experiment 4.

First we compared the proportion of looks to the target and the other four irrelevant objects across the 200-500 and 500-800ms time windows. First t-tests showed that in all six conditions, participants have larger proportion of fixation to the target than the other irrelevant objects in the two time windows (p's < 0.05). Then the proportions of fixations to the target were compared with the chance (one sixth). Proportions of fixations to the target were larger than chance in the second time window across all conditions (p's < 0.01). In particular, in the Non-interactive non-competitor condition: $t=7.708$, $p<0.01$; the Non-interactive competitor condition: $t=5.125$, $p<0.01$. Thus, the bias to target formed at broadly the same rate across all three conditions. This means that the interaction reported above, showing a lack of competitor interference in the Non-interactive conditions was not a result of some kind of floor effect.

Summary of Results:

To sum up the results of the ANOVA analysis, let us first consider the effect involving Blocks found through the two main time windows. In both cases, a Block-by-Competitor type interaction revealed that Non-Competitor trials attracted more Target Advantage than Competitor trials in the First half of the experiment, but there was no difference in the second half of the experiment. This is the kind of pattern we expect in a Perspective-Taking study where participants ‘improve’ performance in the more difficult trials involving a competitor distractor in private positions, relative to the baseline condition involving no competitor.

Let us now turn to effects involving AQ score. Recall that in Experiment 3, we found a surprising interaction between AQ and Question-type revealing that the Low-AQ group had lower TA scores in the baseline Test-question study. The account of that result given above was in terms of implicit vs. explicit social-cognitive ability. Here we found an interaction only in the Preview window showing that in early trials in the study, the more socially able group did not differ between Competitor and baseline Non-competitor trials, while the low social group had lower TA scores when a competitor was in private view. This interaction was not present in the second half of the study. It is difficult to relate this result to that mentioned from Experiment 3.

Next we turn to the Competitor-by-Interactivity interaction that was found in both the 200-500ms and 500-800ms windows. In both cases, when broken down, results revealed that Target Advantage on Competitor trials was lower than the baseline Non-Competitor trials in Interactive and Low-Interactive conditions but there was no difference in the Non-Interactive trial. This result is in line with the predictions outlined at the beginning of the study, that interactivity in relation to the private objects can create interference when a competitor is present. The reason for this is that an interaction between instructor and participant about the private objects makes those object more salient or more available to be considered as potential referents by low-level memory-based processes. When one of those private objects fits the description (‘apple’) that the instructor uses, then it is harder to ignore those objects, compared to the baseline case where no apple is present.

Before resting our conclusions too heavily on these results, we should consider the fact that the basis of these interactions seems to be that Target Advantage on Interactive and Low-Interactive conditions is relatively high, compared to the No-interactive condition, while TA on Competitor conditions seems to be similar across the different sub-experiments. This observation is borne out by post-hoc tests on Non-Competitor and Competitor conditions separately. A three-way ANOVA showed that there was a significant difference in TA for the non-competitor items among three interactivity conditions (200-500ms: $F(2, 59)=4.055$, $p<0.05$; 500-800ms: $F(2,59)=6.721$, $p<0.01$). Follow-up analyses showed that there are larger proportions of looks to the target in the Interactive and Low-interactive condition than the Non-interactive condition (p’s < 0.05), but no difference between the Interactive and Low-interactive condition. By contrast, for the competitor items there was no different proportion of looks to the target among three interactivity conditions (200-500ms: $F(2, 59)=0.850$, $p=0.433$; 500-800ms: $F(2,59)=1.751$, $p=0.183$).

These results tell us that when participants sat in the same room as the instructor and in close proximity to her (see Figure 4.11), their performance on the baseline No-Competitor trials was better than when the instructor was not present in the room at all. This is entirely consistent with our hypothesis since salience or availability of an object for reference depends on the extent to which attention is shared. When the instructor is co-present, there is much more direct evidence available that participants share attention with the instructor to the objects on the screens than when the instructor is absence. For example, the participant sees the co-present instructor is looking at the screen when the common objects appear. Thus they have more situational evidence that attention is shared to the appearance of the common-view objects. This fact in itself should result in the greater salience or availability of the common-view objects for the Interactive and Low-Interactive conditions than the Non-Interactive condition. Thus it should result in the kind of improved TA scores for co-present groups that we see in the Non-Competitor condition. However, it should also lead improved performance in the Competitor condition for the co-present groups. That it does not would be a result of the interaction concerning the private objects. To put this point another way, the co-presence of the instructor on the Interactivity and Low-Interactivity result in a higher baseline Target Advantage score. The experiment was designed to see if the different procedures concerning interactivity (whether instructor and participant share attention to the private objects) would lead to relatively lower Target Advantage score on Competitor trials, relative to the baseline. The prediction was that more interaction about the private objects would lead to a lower Target Advantage, compared to the baseline Non-competitor condition. This is what we found.

The interpretation of these results could be investigated further by the inclusion of another condition where there is interaction about the private objects but the instructor is not co-present. The prediction would be that in this condition, Target Advantage on Non-Competitor trials would be lower than in the Interactive and Low-Interactive (i.e. co-present) conditions here. Relative to the No-Competitor baseline, Target Advantage on Competitor trials in this fourth condition would be lower.

4.3. Discussion

The procedure in this experiment allowed for interaction between the director and participant at a time when the participant was required to pay attention to their private objects. In the Interactive condition, attention toward a property of these objects was shared via a communicative interaction between director and participant. In the Low-Interactive condition, no communicative interaction took place between the director and the participant. However, in this condition, the director was present, sitting close by to the participant and the participant was aware that the director was being informed of the category of the objects at the same time as they were. Thus, there was to some extent, shared attention to a property of the private objects. In contrast, in the Non-interactive condition, the director is not present in the room for the session and the participant is aware that the message about the private objects is not shared with the director. Another feature of the manipulation in this experiment concerns whether the instructor is present in the same room or not. When instructor is present, participants have more situational evidence that she shares attention to both the common objects and the private ones. In the case of the common objects, attention is shared to their identities (whether they are apples, grapes etc.), in the case of private objects, attention is optionally shared to superordinate properties (being fruit, shapes etc.).

Critically, the form of the interaction differed between this experiment and Experiment 3. In the interactive Guess question condition of Experiment 3, the director makes a guess about the category of the participant's objects. Compared to the non-interactive version of that procedure (where participants silently answer a Test question about their category) interactivity improved performance in Experiment 3. We attributed that improvement to the fact that making a guess about objects highlights the director's ignorance of their type rather than simple interactivity itself. In the current experiment, we see in fact that the more interactive conditions yielded a greater change relative to the baseline Non-Competitor condition than a non-interactive condition. Thus, we can conclude that improved performance in Experiment 3 was due to how participants processed the interaction in a way that reinforced ignorance of the type of objects in the participant's private view.

The results of Experiment 4 are in line with the predictions based on my proposals about common ground: when participants were likely to be attending to their private objects, cues to the fact that attention was shared to these objects caused participants to suffer relatively more distractor interference than when there were no cues to interaction at this phase. We attribute this result to the fact that objects to which attention has been shared are common ground objects and such objects become more accessible for definite reference subsequently. Thus participants are more liable to consider such objects, even though they have evidence that the speaker could not know their identities.

An alternative account of our results could point to the fact that in the Interactive condition of Experiment 4, the director's dictating of the superordinate category may create an impression of more knowledge about the objects than the director in fact has. It could thus be that the interaction has an opposite effect to that in Experiment 3 and that is why participants suffer interference. While we cannot rule out that some participants may have been susceptible to this illusion, we point out that in the Low-Interactive condition participants also suffered competitor interference. In the Low-Interactive condition, the director sits passively, like the participant, while the category information appears on the screen. There is no interaction. Thus, any misapprehension about what the director knows about the private objects that is created by the director's behaviour in the Interactive condition of Experiment 4 is unlikely to be created when the director is passive, as in the Low-Interactive condition.

5. General discussion

In Experiment 3 and 4, we explored two levels of cognitive processing in the perspective taking: a lower-level memory bases system for ‘common ground’ based on previous shared experience and a system that integrates information about speaker’s mental states into referential decisions. Experiment 3 demonstrates that more salient information about the director’s ignorance improves the extent of perspective taking. This is the first reported experiment that directly tests the role of Theory of Mind inferences in on-line perspective-taking tasks.

The results from Experiment 4 rule out the possibility that simply the degree of interactivity led to the better performance in the Guess condition of Experiment 3. In Experiment 4, participants in the interactive conditions were less able to ignore the competitor in private view than participants in the non-interactive condition.

The results of Experiment 4 lend support to the proposals about common ground in this thesis. When the director was in the same room, participants were less able to ignore objects in privileged view, because the co-presence and other cues to shared interaction increase the degree to which the existence of private objects are processed as shared attention and thus the degree to which their salience makes them candidates for definite reference.

It may seem surprising that more interactivity between director and participant leads to poorer performance. However, as predicted by the proposals in this thesis, it depends on the nature of the interactivity. When interactivity makes information about the speaker’s mental state more salient, then this can improve performance on perspective taking tasks. When interactivity only serves to increase the shared attention to the existence of private objects, then it makes performance worse.

Our proposals imply that on-line perspective-taking tasks of the form first reported by Kesyar and colleagues can be made more difficult, to some extent, by the operations of the low-level common ground mechanism. We can compare the results from this chapter to those of other results showing that performance on similar tasks is modulated by social factors.

Adult participants in Experiment 4 behaved in a similar way that two-year olds did in the study reported in Moll et al. (2011). In that study 2-year olds were unable to treat as novel the identity of an object for which only cues only to shared attention to the existence of the object were available.[25] In both cases we explain the results in terms of two social-cognitive mechanisms being involved in decisions about definite referents - a lower-level memory bases system for 'common ground' based on previous shared experience and a system that integrates information about speaker's mental states into referential decisions.

Taken together, the results of Experiments 3 and 4 are relevant to how we can interpret results reported in Savitsky et al., (2011) and Wu & Keysar (2007a). For example, Savitsky and colleagues (2011) demonstrated that the familiarity of the interlocutors influenced the establishment of common ground. In that study, participants demonstrated more egocentric behaviours when they perform the game-board task based on Keysar and colleague 2000's study with their friends, compared to strangers. Savitsky and colleagues interpreted this effect in terms of Keysar's Perspective-Adjustment model, using the idea that friends tend to put in less effort to complete the task. While this may be the case, to some extent, it could also be the case that people who are more familiar with each other are better able to, or more liable to, attend to cues to shared interactions. Without any more specific cues to director ignorance, this factor may have led to a greater salience for privately viewable objects when social distance is lower.

[25] This experiment is summarised in Chapter 1.

Similar study on the cultural difference of the perspective-taking found that Chinese people performed better than Western people, which may be due to the fact that in the study Chinese participants were new to the environment (Wu & Keysar, 2007b). There being greater social distance between visiting students from another culture than US-based undergraduate students used in the control group.

The proposal in this thesis is that, in conversation, mechanisms for common ground information and mechanisms for integrating Theory-of-Mind information are separate. As mentioned in Chapter 3, the Perspective-Adjustment model of language processing rests on the idea that some elements of perspective taking are costly, leading to an egocentric – first heuristic. This position is supported by numerous studies indicating a cost factor for perspective-taking. For example, Brown-Schmidt (2009b) demonstrates that performance on perspective-taking task correlates with inhibition control ability. Similarly Lin, Keysar & Epley, (2010) show that working memory plays a big role in the task. Other research, such as Apperly et al., (2006) and Vorauer, Martens, & Sasaki, (2009) speak to a cost for similar tasks. Could it be that these are costs inherent in integrating Theory of Mind inferences into on-line processes? The results in this thesis should make one wary of drawing that conclusion. Here I have argued that, in fact, social cognitive mechanisms for common ground may in some ways make the task more difficult. Working memory and inhibition control may be called on to a greater extent in these tasks precisely because one has to discount salient referents provided by common ground mechanisms. To test this hypothesis, one could design a study to see whether Inhibition control abilities play a greater role in conditions where attention is shared more to the private object.

Contrary to the assumptions behind the Perspective-adjustment model, the results of Experiment 3 demonstrate very rapid integration of Theory of Mind inferences (about the speaker's ignorance). Thus we would be disinclined to align the proposals in this thesis with any apparently similar proposals that compartmentalise abilities for perspective-taking into a costly set and an automatic, low-cost set. As mentioned in Chapter 3, there is evidence that Theory of Mind information can be integrated rapidly in language comprehension (Ferguson & Breheny 2011). We have no reason to believe that there is any in principle delay in integrating Theory of Mind inferences in language processing. Thus, when not coming into conflict, the mechanisms for common ground and for Theory of Mind inferences can in principle have an effect in the same timecourse.

The proposals in this thesis have implications for research into Theory of Mind more generally. This research has often focussed on the false-belief task and found that for children, it is difficult to pass explicit versions before the age of four (Perner 1991; Perner, Leekam, & Wimmer 1987; Wellman, Cross & Watson, 2001) and that for adults the task does pose problems (Apperly et al. 2006). At the same time, infants can pass an implicit version of the false-belief task soon after their first birthday (Onishi & Baillargeon, 2005; Baillargeon et al., 2010; Southgate et al., 2007). Recently it has been demonstrated that 3 year-old children pass explicit versions of the false belief task in Rubio-Fernandez & Geurts (2013). In that study, it is shown that younger children perform better when their perspective is not interrupted by that of the experimenter. It is also shown in Rubio-Fernandez (2015) that adults performance on versions of the false-belief task is subject to interference where perspective in ‘interrupted’.

The proposals in this thesis suggest that one hitherto uninvestigated factor in explicit versions of the false belief task relates to the interference of common ground. In a typical false belief task, the child and an experimenter watch the target object be placed in one location by one character and then it is moved to the so-called ‘true belief’ location. In many such studies, the experimenter demonstrates the story to the child themselves. This form of interaction would strengthen the common ground information shared about the target object. When it comes to testing the child, mention of the object may prime shared information about that object to an extent that it becomes difficult to ignore for younger children. Thus we make a prediction that the extent to which information about the true location is common ground between the child and experimenter when the question is asked will determine the rate of error/interference in this task. We aim to test this hypothesis in future research.

The results reported in this chapter also have implications for the research on the development of children's use of discourse markers. For instance, Matthews et al., (2006) show that 2-year olds but not 3- or 4-year olds have difficulty providing full descriptions of a character in an event to an experimenter who cannot see the event. Since the experimenter is co-present and generally interacts with the child while she watches the events, it may be that the younger children find it more difficult to ignore the cues to shared attention, or to integrate this information with the fact that the experimenter does not know the identities of the character on the video.

Conclusion

The dissertation concerns what cognitive mechanisms play a role in online referential processing and how. In particular, we were interested in how previously shared information and information about interlocutors' mental states are integrated to resolve referential ambiguities in language comprehension. According to Clark & Marshall (1981), common ground is common-knowledge of discourse relevant facts and it serves as the main constraint on definite referential processes. The first chapter argues that, viewed from the perspective on on-line processing, mechanisms for common ground cannot be the main constraint on definite reference, due to the phenomenon of accommodation. Chapter 1 also outlines and motivates an alternative view of common ground. This view is conceptually simpler, since it does not involve common knowledge but rather shared attention. Shared attention has been argued to provide a basis for the specifically 'open' nature of human communication (Barwise, 1988/1989; Peacocke, 2005) and its development has clear links to the development of communicative abilities in pre-linguistic infants (Csibra, 2010; Tomasello et. al., 2005). Although abilities to represent shared attention involve abilities to represent intentions (Peacocke, 2005; Carpenter & Liebal, 2011), it emerges prior to more full-fledged Theory of Mind abilities that manifest themselves in the second year. The proposal about common ground says its operation is supported by lower-level cognitive biases triggered by social cues such as eye-contact, contingent behaviour and self-directed speech. These cues trigger an enhanced processing in terms of greater attention to the shared stimuli and deeper integration in memory. Thus when previously shared information is cued for recall, its salience or availability is greater than if it had not been shared.

In line with the idea of graded salience of common ground (Brown-Schmidt, 2012), the second chapter explored how domain-general memory mechanisms interface with social cognition in the process of making previously shared information available for referential processing.

We tested the memory-based account of shared information due to Horton and colleagues (Horton & Gerrig, 2005). We found evidence consistent with the predictions in this model that low-level or automatic effects of memory associations influence on-line referential processing. However, we also found evidence that memory-based mechanisms are influenced by shared attention at the coding phase.

According to the proposals in this thesis, keeping track of common ground should be low in cost. Yet some studies demonstrate a difficulty in in perspective-taking tasks (for example, Keysar et al. 2000). The difficulty may come from the access to theory of mind, or other task-related factors. The thesis argues that research using the perspective-taking task has not to date specifically examined whether Theory of Mind inferences are integrated in early phases of referential processes. All widely discussed studies do not distinguish between the role of common ground and inferences about the speaker's ignorance of certain objects. So the third experiment intended to test the role of Theory of Mind inferences in perspective taking task.

In the third experiment, participants received a conjecture regarding the category of their privileged objects before the referential task started. There were two types of conjecture, one asked by the speaker, and the other just presented on the screen. Despite only difference in the presentation types, participants less resorted to their privileged ground when the speaker asked the category of the objects, because the speaker's question implied their ignorance of the participants' privileged ground. In contrast, when participants read the conjecture on the screen, they were suffered interference from their privileged competitor objects. If we assume that the procedure in the guessing condition makes the director's ignorance of the private objects' identities more salient than in the non-interactive test condition, then the very early influence of this manipulation (from 300ms after the onset of the critical word) reveals that

Theory of Mind inference can be integrated into referential processes in the perspective-taking task from the earliest stages.

But as discussed, the different interactivity may have had an effect on the results. That is, larger interactivity in the guess condition led to a better performance. So a follow-up study, Experiment 4, examined this possibility.

Experiment 4 was also designed to explore an additional hypothesis that emerged from the proposals in this thesis: that some part of the difficulty in perspective taking studies results from the fact that the existence of privately viewable objects is common ground, even though their specific identities are not. Based on work in the social-cognition literature (Knoblich et. al., 2011) and the developmental literature (Moll et. al., 2011) we conjectured that the co-presence of the speaker increased the degree to which the existence of private objects was shared attention. The fourth study intended to test whether the co-presence of the speaker influenced the perspective taking as predicted, and also provide a baseline for the interactivity of the experiment 3. In the same referential task as experiment 3, the interactivity of speaker was manipulated in three ways: interactive, low-interactive and non-interactive conditions. The interactive and low-interactive condition were in parallel to the guess and test condition in the experiment 3, because they were different in whether the confederate spoke to participants in a pre-test phase. The non-interactive condition was different with the first two on whether the confederate was co-present. The results did not found any effect on the interactive and non-interactive condition, indicating that results of the experiment 3 were less likely due to the interactivity between interlocutors. But it showed that the Interactivity influenced participants' perspective taking.

In sum, the dissertation proposed two relevant levels of social-cognitive processing in the perspective taking task: a lower-level memory bases system for 'common ground' and a system that integrates information about speaker’s mental states into referential decisions. Both mechanisms are separate. The former one relied on automatic memory retrieval of previous shared experience, while the latter one was based on a higher order theory of mind mechanism.

References

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, 17, 841–844.

Aumann, R.J .(1976). Agreeing to disagree. *The Annals of S tatistics*, 4(6) ,1 236-1239.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14, 110-118.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21, 37–46.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders*, *31*(1), 5-17.

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109*(1), 18-40.

Barwise, J, 1988, "Three Views of Common Knowledge", in Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge, M.Y. Vardi (ed.), San Francisco: Morgan Kaufman, pp. 365–379. (reproduced in Barwise, (1989)

Barwise, J. 1989, *The Situation in Logic*, Stanford: Center for the Study of Language and Information.

Bratman, M. (1992) "Shared Cooperative Activity," *The Philosophical Review*, 327-341.

Breheny, R. & Zheng, X. (under review). The roles of Theory of Mind and common ground in perspective taking – a developmental persective. Forthcoming in Hinterwimmer, S., Schumacher, P. & Weiland-Breckle, H. (eds) Special Issue on Perspective Taking. *Lingua*

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.

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of memory and language*, *61*(2), 171-190.

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin and Review, 16*(5), 893-900.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, 27, 62-89.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). Chapter Three-People as Contexts in Conversation. *Psychology of Learning and Motivation*, *62*, 59-99.

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.

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in Another Person's Shoes:Incremental Perspective-taking in Language Processing. *Dialogue and Discourse 2*, 23.

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*.

Clark, H. H. (1992). *Arenas of Language Use*. Chicago, Illinois.: University of Chicago Press.

Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H. H., & Marshall, C. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical issues in natural language processing* (Vol. 2, pp. 57-63). New York, New York.: Association for Computing Machinery.

Clark, H. H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 101 – 131). Cambridge, UK: Cambridge University Press.

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

Csibra, G. & Gergely, G. (2009). Natural Pedagogy. Trends in cognitive sciences. 30(4).

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language* 25: 141-168.

Cubitt, R.P., Sugden, R., (2003). Common knowledge, salience and convention: a reconstruction of David Lewis's game theory. Economics and Philosophy, 19, 175–210.

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology*, *47*(6), 1499-1503.

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.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602-9605.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy*, *5*(1), 39-60.

Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires. *Cognition*,*119*(2), 179-196.

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology*, *48*(1), 257-263.

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, 99–103.

Fodor, J.A. (1998) *Concepts: Where Cognitive Science Went Wrong*. Oxford, OUP.

Frable, D. E. S., & Bem, S. L. (1985). If you are gender-schematic, all members of the opposite sex look alike. *Journal of Personality and Social Psychology*, 49, 459–468.

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

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292-304.

Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, *33*(4), 511-526.

Grice, H. P. (1957). Meaning. *The philosophical review*, 377-388.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 43 – 58). New York: Academic Press.

Grice, H.P. (1982). ‘Meaning Revisited’, in Mutual Knowledge, N.V. Smith (ed), New York: Academic Press, 223–43. Additions 27-07

Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., & Elwell, C. (2008). Early cortical specialization for face-to-face communication in human infants. Proceedings of the Royal Society B: Biological sciences, 275(1653), 2803-2811.

Hallett, P.E. (1986). Eye movements. In K.B. Boff, L. Kaufman & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance I: Sensory Processes and Perception*. pp. 10-102. New York: Wiley.

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.

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters’ thoughts and feelings by able, mentally handicapped and normal children. *Journal of Autism and Developmental Disorders*, 24, 129–154.

Harman, G. (1977). “Review of Linguistic Behavior by Jonathan Bennett”, Language, 53: 417–424.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831-836.

Heim, I. (1982). The Semantics of Definite and Indefinite Noun Phrases. Phd Dissertation, University of Massachusetts, Amherst.

Heim, I. (1983). On the projection problem for presuppositions. *Formal semantics–the essential readings*, 249-260.

Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes, 22*(7), 1114-1139.

Horton, W. S., & 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.

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.

Jespersen, O. (1949). A Modern English Grammar, vol VII. Completed by N. Haisland. London 1961.

Johnson, S. (2003). Detecting agents. Philosophical Transactions of the Royal Society B, 358, 549-559.

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, *4*(1), 71-84.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32-38.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current directions in psychological science*, *7*(2), 46-50.

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.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*(3), 343-346.

Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language, 56*(3), 436-455.

Knoblich, G., S. Butterfill, and N. Sebanz. (2011). Psychological research on joint action: theory and data. In The Psychology of Learning and Motivation, edited by Brian H. Ross, 59-101. *The psychology of learning and theory: Advances in research and theory*. San Diego: Elsevier.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R package version*, *2*(6).

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.

Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". *Psychological review*, *94*(4), 412.

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, *6*(4), 315-324.

Lewis, D. (1969). *Convention: A philosophical study.* Cambridge, MA: Harvard University Press.

Lewis, D. (1994). Reduction of Mind. In S. Guttenplan (ed) *A Companion to the Philosophy of Mind*. Oxford, Blackwell.

Liebal, K., Carpenter, M., & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. Infancy, 15(5), 545–556. doi:10.1111/j.1532-7078.2009.00028.x

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.

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

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, *101*(4), 676.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*(03), 403-422

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.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*(2), 201-213.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1155.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283-312.

Moll, H., & Meltzoff, A. N. (2011). Joint attention as the fundamental basis of perspectives. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 393-413). Cambridge, MA: MIT Press.

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology*, 43(2), 309-317.

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.

Neale, S. (1990). *Descriptions*. Cambridge, Ma. MIT Press

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 8 April, 255-258.

Paternotte, C. (2011). Being realistic about common knowledge: a Lewisian approach. *Synthese 183.2* , 249-276.

Peacocke, C. (2005). Joint attention: Its nature, reflexivity, and relation to common knowledge.

Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

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.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 169-190.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347.

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

Rawson, K. A. (2004). Exploring automaticity in text processing: Syntactic ambiguity as a test case. *Cognitive Psychology*, *49*(4), 333-369.

Rubio-Fernández, B. Geurts. (2013). How to pass the false-belief task before your 4th birthday Psychological Science, 24 (1), pp. 27–33.

Russell, B. (1905) On Denoting. *Mind*. Vol. 14, No. 56 pp. 479-493

Russell, B. (1919). *Introduction to Mathematical Philosophy*. London.

Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, *47*(1), 269-273.

Schiffer, S. R. (1972). *Meaning*. Oxford: Oxford University Press.

Schneider, D., Slaughter, V., Bayliss, A. & Dux, P. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, 129(2): 410-417.

Skyrms, Brian. 1996. *Evolution of the social contract*. Cambridge University Press

Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. Developmental Psychology, 44, 1789-1795.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587-592.

Southgate, V., Chevallier, C., & Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science*, 16, 907-912

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 16, 907-912.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.

Stalnaker, R. (1974). Pragmatic Presuppositions. Semantics and Philosophy, ed. by MK Munitz and P. Unger, 197-213. *New York: University Press. Reprinted in Pragmatics. A Reader, ed. Davis, S*, *1991*, 471-482.

Stalnaker, R. (1975). Indicative Conditionals. Philosophia, 5: 269–286.

Stalnaker, R. C. (1978) Assertion. In: *Syntax and semantics, vol. 9: Pragmatics,* ed.P. Cole, pp. 315–32. Academic Press.

Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, *25*(5), 701-721.

Symeonidou, I., Dumontheil, I., Chow, W-Y., & Breheny, R. (*forthcoming*) Development of online use of theory of mind in adolescence: an eye-tracking study. *Journal of Experimental Child Psychology*.

Taylor, S. E., & Falcone, H. T. (1982). Cognitive bases of stereotyping: The relationship between categorization and prejudice. *Personality and Social Psychology Bulletin*, 8, 426–432.

Thomason, R., (1990) Accommodation, meaning, and implicature: interdisciplinary foundations for pragmatics." In, *Intentions in communication*, P. Cohen, J. Morgan, and M. Pollack, eds., MIT Press,

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675 - 691.

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what's new for other persons. *Developmental Psychology*, 39, 906–912.

van Twuyer, M., & van Knippenberg, A. (1995). Social categorization as a function of priming. *European Journal of Social Psychology*, 25, 695–701.

Vorauer, J. D., Martens, V., & Sasaki, S. J. (2009). When trying to understand detracts from trying to behave: effects of perspective taking in intergroup interaction. *Journal of personality and social psychology*, *96*(4), 811.

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

Wu, S., & Keysar, B. (2007a). The effect of information overlap on communication effectiveness. *Cognitive Science, 31*(1), 169-181.

Wu, S., & Keysar, B. (2007b). The effect of culture on perspective taking.*Psychological science*, *18*(7), 600-606.

Zheng, X. & Breheny, R. (under review). Constraints on memory-based mechanisms for common ground. *Journal of Experimental Social Psychology.*

Zheng, X. & Breheny, R. (in prep). Listeners' eyes reveal separate roles for Theory of Mind and common ground in on-line perspective taking. Journal of Experimental Psychology: LMC.

Note:

The present thesis draws on three papers that are either under review or in prep now. Chapter 2 draws on Zheng & Breheny (under review). Chapter 3 draws on Zheng & Breheny (in prep). Chapter 1 draws on Breheny & Zheng (under review).