Multimodal and Multisensory Characteristics of Deictic Communication

Gozdem Arikan

100290988

School of Psychology

University of East Anglia

Primary Supervisor: Prof. Kenny R. Coventry

Secondary Supervisor: Dr. Louis Renoult

30 September 2023

Thesis written to fulfil requirements for the degree of Doctor of Philosophy.

Acknowledgements

One thing a PhD candidate thinks about throughout their academic journey is the acknowledgement section of their thesis. Moments come in your journey where you think - this could make it into my acknowledgements! These are the moments that leave the strongest remark on this path, not only for mere completion of your PhD, but moments of gratitude reflected in one's future career. Writing this section gives an opportunity to show gratitude to both academic and non-academic supporting networks of people. In all honesty, the gratitude felt by myself for people I mention below is so strong that I thought of not including this section altogether - maybe also because I do not believe in or like goodbyes.

I would like to start this expression of sentiment with my friends and family, with specific shoutouts to some names (hoping that ones I have missed understand that their support was not less than those mentioned). First of all, I would like to thank Mustafa Tekman and Dr. Laszlo Horvat for their support, both emotionally and intellectually. I would then like to thank my PhD cohort, "awesome support bubble", Agatha Bataille, Laura Manno and Alice Bush, with a specific shoutout to my dearest colleague and friends Paula Janjic and Greta Melega. Without their input and support, my PhD journey would not been half as enjoyable. I would also like to thank my best friend Alya Basak (soon to be Dr!) for being the one person on this planet (possible in all alternate universes) to share the same vision of life and career, from pretty much the very beginning of our existence. I would also like to thank Dr. Prerna Aneja and Dr. Alvaro Rodriguez for our intense discussions over a glass (or two) of wine, forming the basis for lifetime friendship. I would also like to thank my dearest friend Dr. Katja Waschneck, with whom I spent entirety of the Covid-19 pandemic lockdown in Norwich, and I couldn't have asked for a better company during this period.

I would like to thank my past and present supervisors, and people whom I can hope will be my future collaborators in the years to come. Firstly, I would like to thank Prof. Albert Postma, who gave me an opportunity during my master's internship to delve into the world of language and perception.

Secondly, I would like to thank my secondary supervisor, Dr. Louis Renoult for helping me analyse EEG data (not reported here, sorry!) and come with creative ideas to approach my topic. Lastly, and with the highest gratitude, I would like to thank my primary supervisor Prof. Kenny R. Coventry. He has not only been a supervisor, guiding me on the right track for my research but also the best mentor I could ask for, supporting me in becoming a well-rounded researcher. He pushed me to the limits of my ability, and listened and understood all my ideas and thoughts from the time we met during my masters. In my future endeavours, I count on our new research collaborations, working on topics we initially got inspired to work on (which unfortunately got stunted due to COVID-19; but even then, he had ideas to keep me busy!). There are not enough words to thank him, and his consistent leadership.

I would like to thank Dr. Peter Boddy, who very quickly adapted to working together, although miles apart. Not only did we work together to create a number of online experiments, but I also learnt a great deal from him as a young researcher. I would also like to thank my fiancé Joeri van Helden (also soon to be Dr.!), not only for his consistent emotional support, but also for modelling in our kitchen to become part of the stimuli for my research! I would also like to thank Joe Murphy who helped me collect data in our collaboration with Vision Norfolk. With this remark, I would like to thank staff and members of Vision Norfolk, especially Edward Bates, for help in recruiting blind and visually impaired individuals, and opening the doors of their Great Yarmouth and Norwich hubs for data collection. Speaking of running experiments, I cannot finish this acknowledgement section without thanking the tech team of school of psychology, specifically Richard Carey-Knight, Lizzie Watson and Malcolm Rae. Each contributed to my research dearly, even when I needed the longest cables and 3D printed objects - much appreciated!

I would lastly like to thank University of East Anglia, School of Psychology and Graduate school for funding my studentship and providing me with resources to conduct the research reported below.

Author's Declaration

This PhD was funded through a studentship provided by the University of East Anglia, School of Psychology. All the experiments were approved by University of East Anglia School of Psychology Ethics Committee. The work from this PhD was presented at international and multidisciplinary conferences and submitted for publication to journals listed below. The experiments submitted for publication are presented in the thesis in manuscript format.

Conferences:

Arikan, G., Boddy, P., Coventry, K (26-29 July 2021). Action Speak Louder than Words and Gaze: The Relative Importance of Modalities in Deictic Reference [Poster presentation]. *The 43rd Annual meeting of the Cognitive Sciences Society*, Virtual event.

Arikan, G., Boddy, P., Coventry, K (1-4 August 2021) Relative importance of modalities in deictic reference [Talk]. *Spatial Cognition 2020*/1. Latvia, Hybrid event.

Arikan, G., Postma, A., Renoult, L., Coventry, K (22-15 May 2022). Deixis without sight: Use of spatial demonstratives in the absence of vision [Poster Presentation]. *Eurohaptics 2022*, Hamburg.

Arikan, G., Boddy, P., Coventry, K (13-15 September). Understanding the Multimodal Nature of Deictic Reference [Talk]. *GeSpin, 2023*, Nijmegen.

Manuscripts submitted:

Experiment 1 submitted to *Psychonomic Bulletin & Review*.

Arikan, G., Boddy, P., Coventry, K (under review). What's the point? Understanding the relative importance of object position, gesture, and eye gaze on spatial demonstrative choices (under review).

Experiments 3 and 4 to *Journal of Experimental Psychology: Learning, Memory and Cognition*:

Arikan, G., Boddy, P., Coventry, K., (under revision). Understanding deictic communication: relative importance of pointing, eye gaze and linguistic cues.

Experiment 5 to *Cognition*:

Arikan, G., Postma, A., Renoult, L., Coventry, K., (under review after revision). Spatial Demonstratives and Multisensory Space.

Abstract

When people communicate, they use a combination of modalities - speech, gesture and eye gaze - to engage and transmit information to an addressee. Deictic communication is a paradigmatic case, with spatial demonstratives (*this*/*that*) frequently co-occurring with eye gaze and pointing gestures to draw the attention of an addressee to an object location (e.g., *this cup*; *that chair*). The use of deictic expressions can be influenced by range of spatial, perceptual, cognitive and social parameters. This PhD Thesis has two foci: the multimodal characteristics of deictic communication, and the implications of multisensory perception on demonstrative choices. In relation to the first focus, four online experiments are presented which investigated the relative importance of spatial demonstratives, pointing and eye gaze as deictic expressions. The results from these experiments overall suggest a dominant effect of pointing gesture in cueing object position, with demonstratives as a relatively weaker means of deictic reference. The nuanced effects and interactions between different modalities are discussed. In relation to the second focus, the notion that spatial perception is a determinant of demonstratives (peripersonal versus extrapersoanl space) is considered in the context of the multisensory nature of peripersonal space. Using a haptic adaptation for a previously validated methodology eliciting spatial demonstrative use (Coventry et al., 2004), in two experiments, we look at the implications of sensory domain(s) used for spatial perception on demonstrative use, testing use of demonstratives when participants can see, feel, or see and feel objects prior to referring to them. Results, including some testing of visually impaired individuals, shows that the effect of referent distance on demonstrative use remains stable irrespective of the modality used to experience space. In the final chapter, general implications of the findings and ideas for future research are discussed.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate ‘take down’ action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents:

List of figures

List of Tables

Chapter 1 General Introduction

In our daily lives, we use a range of expressions to communicate with each other. Words, bodily movements and even where we look as we talk inform an addressee about our thoughts and intentions. The topics of our conversations are shaped by what we see, hear, feel or taste. Something as simple as asking for your favourite cup to drink your coffee from will include a vast number of environmental, cognitive, psychological and linguistic processes. In many ways, deictic communication - the focus of this thesis - is prototypical of these processes, with words (spatial demonstratives: *this cup, that saucer*), gesture and eye gaze working in concert to communicate information about object location.

In this chapter, we will first review the literature on spatial demonstratives, deictic communication and its multimodal nature. We will then look at one of the key factors underpinning demonstrative use, that is (peripersonal versus extrapersonal) space, from a multisensory perspective. This chapter sets the stage for a programme of experiments delving deeper into the multimodal and multisensory characteristics of deictic communication.

1.1 Deictic communication

The term "deixis" (Ancient Greek noun "δεῖξῐς", to point , demonstrate, to show, point out, indicate) refers to the context-dependent use of words and non-verbal expressions (such as pointing and eye gaze) to establish reference. A range of words have deictic functions, including demonstratives (*this*, *that*), pronouns (*he*, *she*, *it*) and adverbs (*there*, *here*, *then*, *now*). Levinson (1983) defines deixis as "the ways in which language encode or grammaticalize features of the context of utterance or speech event" (pp.54), arguing how deictic expressions, such as spatial demonstratives, together with gestures (also termed 'indexical signs'), only have a meaning in a specific context.

1.2 Spatial demonstratives

According to Diessel (1999; 2006), demonstratives are one of the oldest word classes to emerge in the evolution of grammar. They are one of the most used words in English (and across all languages), positioned in the top 20 words in terms of word frequency (Levinson, 2018; Pagel et al., 2013). Demonstratives can be used to describe distance, indicating the position of an object within context relative to the deictic centre, both simultaneously and from memory (Diessel, 2014). Similarly, demonstrative words are used to establish joint attention (Diessel, 2006), engage a listener, and the draw attention of interlocutors to a specific position in space with more of an exophoric usage (Burenhult, 2008; Louwerse & Bangerter, 2005). They can also be used to refer to a point in time or an item in a discourse (in speech and written) with more of an endophoric function (Diessel 1999; Filmore,1997; Levinson, 1983; Lyons, 1977; Maes, Krahmer & Peeters, 2022). Although traditionally argued to be distinct functions of linguistic reference, Talmy (2020) suggests that anaphoric (referring to an element in discourse) and deictic (referring with spatiotemporal function) use of demonstratives share the same cognitive system. The demonstrative *that* is also suggested to have a definite function similar to the word *the* (Ionin et al 2012; also see Ahn 2017; Roberts 2002; Wolter 2005).

The use of spatial demonstratives is thought to be universal (Diessel, 2006; Kemp, 2007). Every language has a demonstrative system, although the number of available demonstrative words and their use vary across languages (Burenhult, 2008). However, how demonstratives are used between and within languages is subject to debate. A wide range of factors influence the way demonstratives are used, including the position of the interlocutors relative to each other and referents, the properties of the objects, and shared attentional space. For example, while English has a two-word demonstrative system (*this* and *that*), Spanish, Japanese and Turkish demonstrative systems have three words (Rubio-Fernandez, 2022). In English, the demonstrative *this* is generally used to refer to objects within reach and *that* for objects outside reach (factors influencing demonstratives will be elaborated later in this chapter). However, the view that the use of spatial demonstratives is only based on spatial cognition has been challenged by some cross-linguistic studies (Diessel, 2014; Evans & Levinson, 2009; Kita & Özyürek,2003; Skilton & Peeters, 2021). It has been argued that the egocentric proximal-distal distinction (relative perception of closeness) of the English demonstratives might not apply in other two-system languages where the meaning of the demonstratives does not solely rely on the referent's position (Enfield, 2003). To give another example, in Turkish (a three-term system), the demonstrative *su* is used to reorient the addressee's attention when it is diverted (Özyürek, 1998; Küntay & Özyürek, 2016), a linguistic function not bound to the spatial position of referents (I will further discuss the cross-linguistic variations in demonstrative use in Chapter 5).

Spatial demonstratives are among the first words children learn, although there is debate regarding the exact age of their acquisition and comprehension. While some research suggested that they are among the earliest words infants learn (Clark, 1978), recent evidence indicates that the use of demonstratives continues to develop over several years. According to Diessel and Monakhov (2022), English-speaking children start using demonstratives between 12 to 17 months old, and they use demonstratives with their spatial function frequently, although the use of demonstratives decreases with age as the children's lexicon expands. On the other hand, Clark and Sengul (1973) argue that preschool children do not start to fully comprehend the distal-proximal distinction of demonstratives. Although they use these words, preschool children do not understand the difference between *this* and *that* to use them contrastively and children are also biased towards a nearby referent (also see Todisco et al., 2020). Furthermore, according to Gonzalez-Pena et al. (2020), children rely more on gestures than demonstratives in the early development of deixis and for the learners of English language, demonstratives do not emerge until the age of two while Spanish-speaking children start using demonstratives at 18 months. Guijarro-Fuentes et al. (2022) also have shown that, although children aged between 2 to 10 years have demonstratives in their lexicon, they do not use the distal term often and their use of demonstratives only fully develops by age 10 (in Spanish). Nonetheless, in relation to other word classes, the first emergence of demonstratives is undeniably early.

It is now established that a number of factors influence demonstrative choices. These factors can be spatial and geographic as well as semantic, psychological and social. A wide range of research either pitted potential factors influencing demonstrative choices against each other or presented the determinants of reference in a complementary fashion. Recently, more holistic views on what shapes demonstrative use have been presented (such as Coventry et al., 2023;

Diessel & Coventry, 2021; Peeters et al., 2020; Terenghi, 2020). In what comes, we will have a brief look at some of the determinants of demonstratives with a specific focus on referent proximity. Although a range of other factors is argued to influence the demonstrative choices, the referent's perceived closeness tends to be a strong and consistent factor across languages, although the effects of non-spatial and spatial factors are not mutually exclusive (Diessel & Coventry. 2021).

1.2.1 Spatial perception and demonstratives

An abundance of experimental and observational studies suggest that referent proximity is an important factor in demonstrative choices (Wilkins, 1999 but also see Kemmerer, 1999). Coventry et al. (2008; 2014) presented experimental evidence in English and Spanish showing that objects within reach, also known as peripersonal space, are more likely to be referred to with the demonstrative *this* (*este* in Spanish) and objects out of reach, in extrapersonal space, with the word *that* (*ese/aquel* in Spanish). Therefore, peripersonal space is an important factor in demonstrative choice. Similar findings have been found in many other languages, also for languages with demonstrative systems consisting of two or more terms (Coventry et al., in press; Meira & Terrill; 2005; Piwek et al., 2008). Other spatial factors influencing demonstratives include the degree of elevation of the speaker and referent (Diessel 1999; Forker, 2020; Peeters et al., 2021) and the height of the referent (Anderson & Keenen 1985; Dixon 2003). Other non-spatial yet physical factors influencing demonstratives include object visibility and accessibility (Coventry et al., 2014; Jarbou, 2010).

In what comes, I will briefly describe the memory game paradigm designed by Coventry and colleagues- an experimental protocol to study demonstrative choices. This methodology is one of the only experimental methods to systematically vary referent (and interlocutor) position and referent characteristics such as familiarity, ownership and accessibility to investigate determinants of demonstratives.

The memory game paradigm was designed to measure determinants of demonstratives in an experimental setting. Variables which had been manipulated through this paradigm include referent position, the position of interlocutors, referent familiarity, visibility, ownership, reachability, and the person placing the referent on their designated position (either the participant or experimenter-also referred as the "who-places" effect) (Caldano & Coventry, 2019; Coventry et al., 2008). This paradigm was designed to measure demonstrative choices and it has also been used to test the effect of demonstrative form on memory for object location (Gudde, et al 2015; 2018). This methodology has been used across different languages including Dutch, Ticuna, Latvian, Italian, Spanish and most recently across 29 languages (Coventry et al., in press). In the basic format of this method, a speaker (the participant) points towards a referent and utters a sentence using a spatial demonstrative, adjective and name of the shape/object (the three-word rule: e.g., *this/that red square*). The objects vary in their position, presented one object at a time.

To avoid any priming in this paradigm testing demonstrative choices, participants are told that they are taking part in an experiment looking at the effects of language on memory, and they are in the language condition. Participants are told that they must use the three-word rule to label the object so that all participants use the same amount of language (to describe the location of an object placed at various positions in front of them on a large table: Figure 1.1). To maintain the memory cover throughout the task, the participants are asked to indicate where a previously placed object was located as accurately as possible (see also Gudde et al., 2016 for memory variation of the task).

Figure 1.1: Schematic representation of memory game paradigm.

a) A referent appears in one of the possible 12 object positions, one item at a time. The participant is seated behind a 320 x 160 cm table on which objects were positioned equidistant (split to 25 cm starting from the participant, b) Example of referents used in early variations of the memory game paradigm, comprising discs with different shapes and colours.

Coventry et al. (2008) aimed to observe demonstrative choices (in English and Spanish) of participants as a function of referent position (individual differences in reaching capacity were accounted for using *post hoc* arm measurements at the end of the experiment), the person placing a referent to its designated position on a given trial, and where the participant (speaker) and addressee (experimenter) were seated (either side-by-side or opposite each other at either end of the long table on which objects were placed). Participants were given instruction cards (36 instructions across 36 trials) which stated the target referent (Figure 1.1b), the position of the referent (one of the 12 coloured dots, as in Figure 1.1a) and the person to place the referent to the designated position (the instruction would read, for example, "I place red square on green dot"). After the object was placed and interlocutors returned to their positions, the participant would then point (either with a stretched arm or a 70cm long stick in this particular study) and refer to the referent with the three-word rule. Results from this study showed an effect of object position, actor, and tool use both in English and Spanish demonstrative systems. When the object was placed within the reach of the speakers of the two languages, the participants were more likely to use a proximal demonstrative (*this*/*este*). Tool use also expanded the peripersonal space with increased use of the proximal demonstrative beyond the end of the finger to the end of the stick. There was also a "who places" effect; when the participants placed the referent, they were more likely to use the proximal demonstrative. Additionally, there was an effect of the addressee position in Spanish and this effect was reliable even with tool use. In the Spanish-speaking group, participants were more likely to use the proximal demonstrative in the second quadrant in condition where the experimenter was facing the participant compared to the condition where the two interlocutors were sitting side by side.

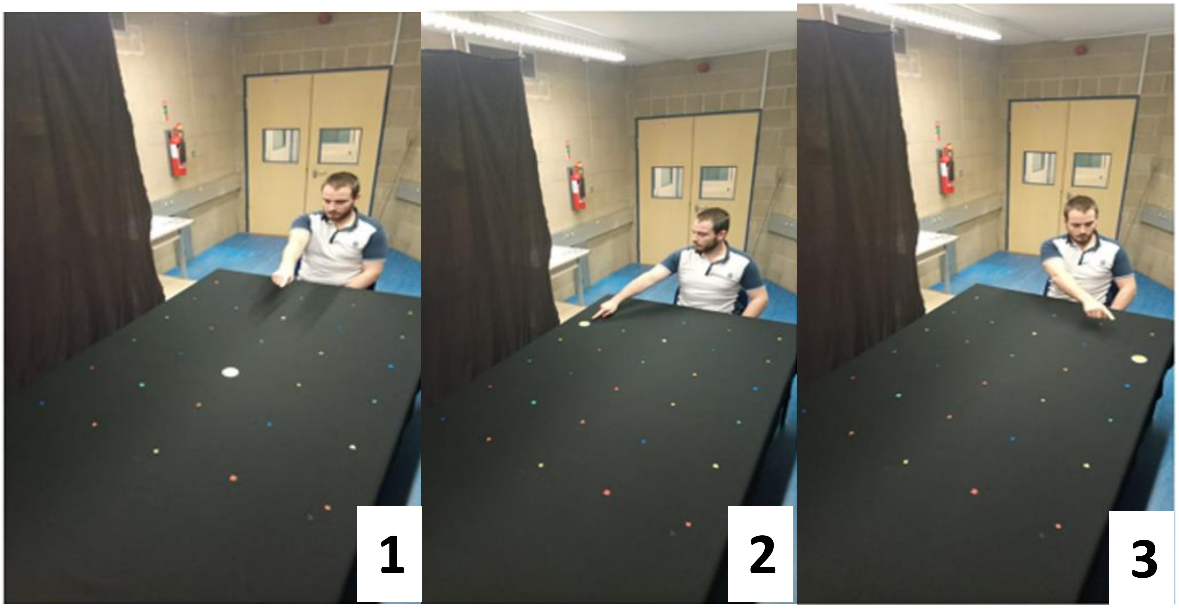
In Coventry et al., (2014), further variables such as object ownership, visibility and familiarity were identified as factors influencing demonstrative choices. Overall, when the participants owned the object, they were more likely to refer to it with the word *this*. The percentage use of *this* decreased as the referents were occluded (covered with a metal cup) compared to when they were the same distance away but visible (uncovered or covered with a glass cup). Finally, familiar objects (familiar colour shape combinations, e.g., white cross) were referred to with the word *this* more often than unfamiliar objects (e.g., less familiar colour and shape combinations, e.g., viridian nonagon).

In another study using this paradigm, Caldano and Coventry (2019) expanded the object positions laterally and shortened them on the sagittal plane with the aim of observing the effect of reachability on demonstrative choices. Laterality was manipulated as a measure of reachability where the referent was within reach with one hand (e.g. the left) and out of reach with the other (e.g. right hand). The experimenter placed discs in one of 30 possible positions. Participants then followed the standard three-word rule producing sentences, such as *this red disk*, as they pointed to the object with their left or right hands (there was no effect of handedness). Results supported the near-far distinction of demonstrative choices also on the lateral plane, with the word *this* used more often in near lateral positions than farther positions (Figure 1.2.a). Furthermore, there was an effect of hand used to point. When participants were pointing with the hand on the same side as the target object (ipsilateral), they used the word *this* more often compared to trials where the object's lateral alignment was on the opposite side of the pointing hand (contralateral), for example pointing to a target on the left side with the right hand (as in Figure 5b second image compared to third image).

Figure 1.2: Lateral layout of Caldano & Coventry (2019).

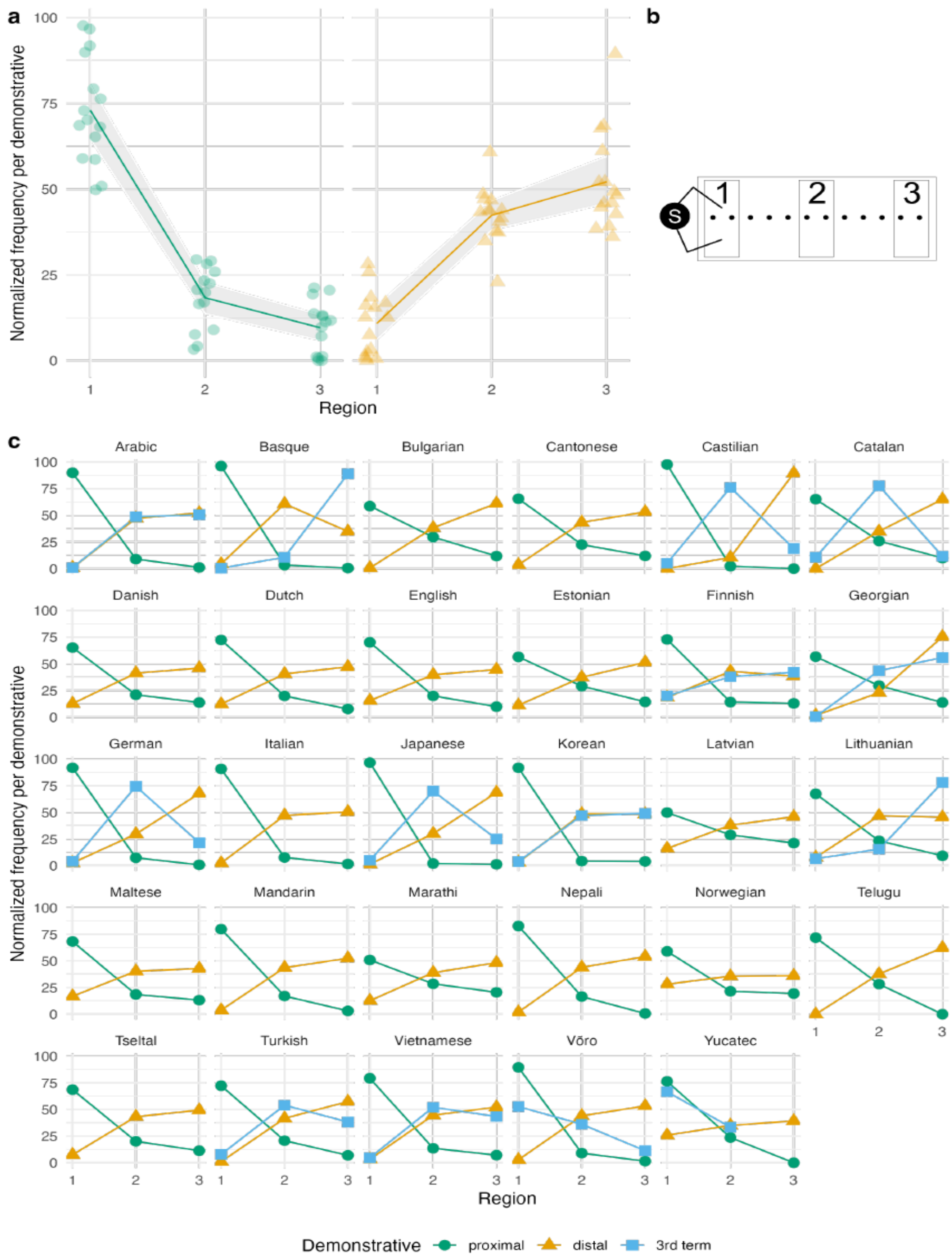
a) Schematic representation of potential referent positions. The object positions varied on both lateral and sagittal planes. Numbers represent the sequence of objects. b) (1) Sagittal pointing, (2) ipsilateral pointing, (3) contralateral pointing based on hand used and position of referent

b



In a recent study by Coventry and colleagues (in press), the memory game paradigm was used to test the influence of object distance and addressee position by speakers of 29 different languages. In this variation, the experimenter (also "the addressee" for the act of referring) places objects to one of the possible object positions and the participants refer to these targets with the three-word-rule mentioned earlier. The authors divide the 12 positions into three regions, with the 3rd region being farthest from the speaker (Figure 1.3.b). They also varied the position of the addressee, who was either next to the participant or facing the participant at the opposite side of the table. Crucially, some of the languages tested have demonstrative systems with more than two words. As we can see from Figure 1.3.a., overall, all languages used proximal and distal demonstratives in their respective languages with proximal demonstratives being used in proximal positions (region 1) and the frequency of proximal demonstrative use reduced as the referent was placed farther, with the distal term used for locations outside of reach. We can see such a trend for proximal terms in all of the languages individually as well (Figure 1.3.c). The use of the third demonstrative seems to vary across languages but based on the visual observation of the data in Figure 1.3.c, participants were more likely to use it in regions outside peripersonal space (outside Region 1). Additionally, there was an effect of addressee position in 8 of the languages tested (some three-term and some two-term languages). These results support the universality of proximal-distal distinction in demonstrative systems across languages.

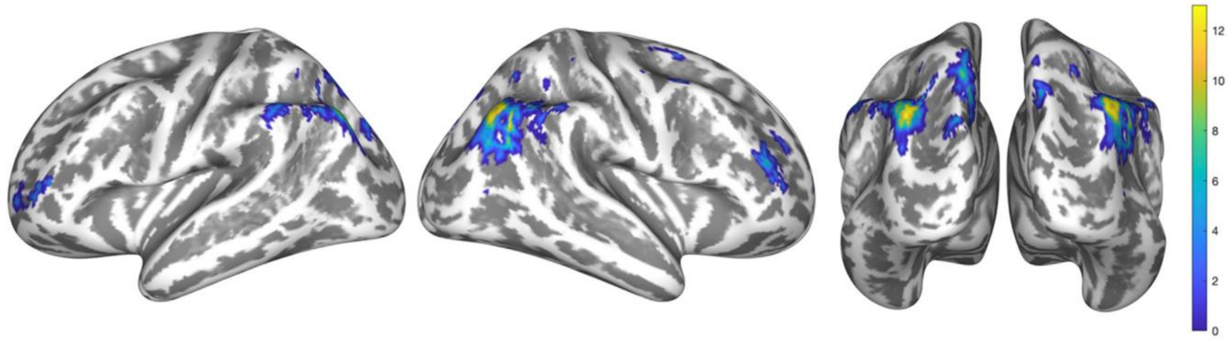
Figure 1.3: Frequency distributions of demonstrative forms in the overall data set and individual languages from Coventry et al. (in press)



Overall, the results of the series of memory game studies provided compelling evidence for a mapping between demonstratives and reachable/non-reachable space, consistent with some earlier accounts of demonstratives that have argued for such a mapping (e.g., Diessel,2006). Moreover, the case for spatial demonstratives mapping onto space is further bolstered with results of a recent neuroimaging study by Rocca et al. (2020). Findings of Rocca et al. (2020) indicate the involvement of brain regions reflecting the role of visuospatial processing, attention regulation and location processing when processing demonstratives. In an fMRI study, participants listened to a naturalistic dialog which had demonstratives integrated into the topic of discussion. Researchers looked for the hemodynamic changes as a response to hearing these words. Results suggested that the parietal (areas associated with visuospatial functions) and the frontal (areas involved in attention regulation) areas and the dorsal stream (region associated with processing object location) were active in processing demonstratives (Figure 1.4).

Figure 1.4: Brain regions responding to spatial demonstratives at Rocca et al. (2020).

The analysis revealed "significant clusters in the inferior parietal cortices, in the medial superior parietal cortices as well as in the middle frontal gyri and right frontal eye field" (Rocca et al., 2020, pp.9).

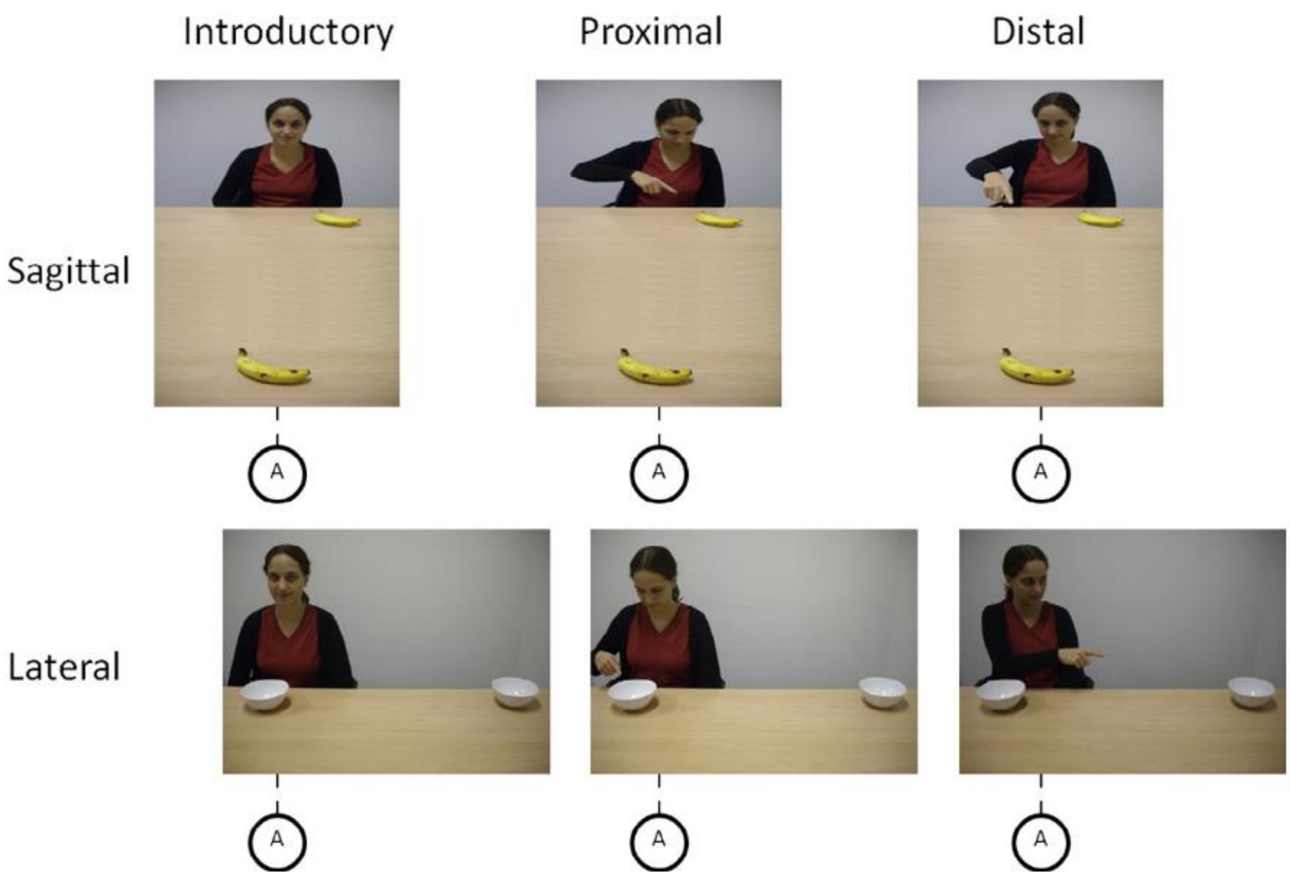


1.2.2 Other determinants of demonstratives

A rather different theoretical approach to demonstratives that has often been considered as a counterargument to the role of referent proximity to the speaker, assumes a listener-anchored or dyadic approach to demonstrative use. In this approach, the relative positions of the speaker and addressee are critical, distinguishing between speaker and listener-anchored demonstratives (Hyslop 1993 as referenced in Burenhult, 2008). Jungbluth (2003) argued that both speaker and listener contribute to demonstrative choices and a common, shared space is established between the interlocutors where the addressee is as important as the speaker in determining the appropriate demonstrative to utter. Peeters, Azar & Özyürek, (2014) further distinguish between the function of proximal and distal demonstratives in relation to the position of interlocutors. They suggested an effect of distance when the referent was near the speaker where proximal demonstratives were used more often. However, the use of distal demonstratives relied on the established joint attention between the listener and the speaker. They argued that while the use of proximal demonstrative is speaker-anchored, distals can also be hearer-anchored. In an EEG study with two experiments, Peeters, Hagoort and Özyürek (2015) further propose that the shared/aligned space between interlocutors is a more important factor than the physical proximity. They carried out an EEG experiment which coupled images (of an individual pointing and gazing towards a target referent) with auditory phrases such as “I have bought this plate at the market” (in Dutch). They manipulated the congruence of the image-audio coupling. The orientation of the objects was also a factor, presented either sagittal or laterally (Figure 1.5).

Figure 1.5: Stimuli example used by Peeters, Hagoort and Özyürek (2015)

The participants saw images of a person referring to a target either near or far from the deictic centre. The referents were positioned either on the sagittal or lateral plane. The images were coupled with audio of sentences consisting of a demonstrative e.g., I bought this banana from the market.

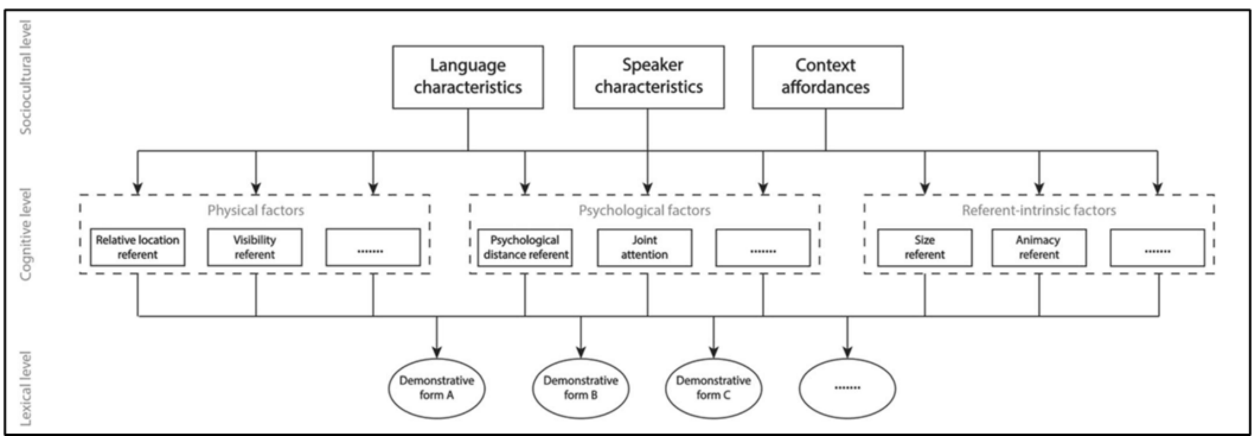


In the first experiment, the results suggested no significant interaction of object distance and demonstrative heard while the orientation had a significant effect (although, curiously, the pre-test judgements of the task by participants piloting the stimuli confirmed a mapping between demonstratives and physical proximity). When the objects were placed on the sagittal plane, there was more negative wave when participants heard a distal demonstrative instead of a proximal, and this was not related to the referent's proximity to the speaker. On the lateral plane, there was no difference in signals for the two demonstrative forms. Overall, the authors argue these results to support the dyad-oriented approach of demonstrative use. For the second experiment, they slightly adapted the images where the objects were more speaker-centred and the distance between potential referents was increased. After the experimental adjustments, the results further confirmed and refined a dyad-oriented account for processing spatial demonstratives. Based on the results, particularly for the images with referents on the sagittal plane, authors argue that, when interlocutors' viewpoints and body orientations are aligned, they create a shared space and when the referent was within this shared space, participants showed preference for proximal demonstratives (and preference of distal demonstrative when outside this space). Furthermore, when there were two potential referents, one within and one outside of the shared space, participants would show a preference for the distal demonstrative regardless of the objects' physical proximity to the speaker. In a review paper, Peeters and Özyürek (2016) further argue for a "collaborative enterprise" of spatial demonstrative use where the speaker and the addressee interact to formulate a shared space onto which deixis lies.

The choice of demonstratives is also argued to be associated with the referent semantically. Rocca et al. (2019) suggested that the referent's (animate and inanimate objects) function and properties might influence demonstrative choices (alongside other factors). They suggest that objects which are easier and safer to manipulate were more likely to be labelled with proximal demonstratives (in English, Danish and Italian. Also see Todisco, Rocca & Wallentin., 2021). Another non-spatial factor which influences demonstrative choices is object ownership (Coventry et al., 2014). For example, in Coventry et al (2014) participants were asked to refer to objects in varying positions. When the referent belonged to the participants (speaker) they were more likely to pick the proximal demonstrative when they referred to the object compared to the referent which belonged to the experimenter (addressee).

Bringing together the wide literature on the determinants of demonstratives, Peeters et al. (2021) developed a conceptual framework for studying and understanding deictic reference (see also Terenghi, 2020). This framework categorised determinants of demonstrative under three major levels of factors: lexical, cognitive and sociocultural (Figure 1.6).

Figure 1.6: Framework to study deictic reference (spoken) by Peeters et al. (2021).
The framework implies three levels of factors influencing deictic expression as a top-down process. At the bottom level we have lexical characteristics of the language. The mid-level considers cognitive processes. The top level named the sociocultural level considers person and context-specific moderators of demonstrative choices.



The first level regards the lexical characteristics of the demonstrative system, such as different forms of demonstratives in a specific language. At the second, the cognitive level, there is a range of factors that include not only the physically perceived position of the referent but also factors such as referent characteristics and shared attention between interlocutors. Peeters et al. (2021) argue that the proximal-distal distinction of demonstrative choice might be "too simplistic”, implying involvement of other cognitive processes than spatial cognition alone and suggest three categories of cognitive processes interplaying in choice of demonstratives: physical, psychological and referent-intrinsic. At the cognitive level the physical factors include the relative distance of the referent from the speaker and the addressee (within peripersonal space or not), referring to the experimental work of Coventry and colleagues in general. Peeters et al. (2021) then argue that the effect of proximity might be context-dependent. In the physical factors, they also include object visibility while acknowledging that it may not be a universal factor.

The second category within the cognitive level, psychological influences, then includes whether the speaker and listener are within the same shared interactional space where the speaker takes the cognitive status of the listener into account e.g., by establishing joint attention through gaze and gesture, to make sure that the referent is cognitively accessible to the speaker. This factor encompasses the earlier work of Özyürek and Peeters, with a dyadic, socio-centric approach to demonstrative use. In the psychological category of the cognitive level, the authors also include the affect towards the referent (termed as “psychological proximity”). For example, a referent with negative valence might be referred with a distal demonstrative. The third category within the cognitive level considers the referent-intrinsic factors including the grammatical associations of the referent such as referent form, properties and associated semantics such as grammatical gender, object animacy, size, harmfulness and familiarity.

At the third and final level of the framework, termed the sociocultural level, the authors argue for three further categories of determinants: language characteristics, speaker characteristics and context affordances. In the language characteristics level, they situate factors which influence some language forms but not others (Diessel, 1999) which are different from the lexical level of demonstrative words (e.g., this level does not include the number of demonstratives available in the demonstrative system of a language). The factors grouped as the speaker characteristics concern the personal characteristics of the speaker including their theory-of-mind development, and individual differences (both between and within cultures). And finally, context affordances refer to context-dependent factors such as the accessibility of referent, the presence of more than one referent which would increase cognitive load/compete for attention. The factors within the three levels of the framework would interplay in a top-down fashion, influencing and formulating the deictic expressions.

To sum up, demonstrative use is determined by an interplay of physical-perceptual, psychological and social processes. The spatial and non-spatial approaches to demonstratives are compatible and work in a harmony of cognitive, psychological, social and linguistic processes formulating deixis, based on the function and context of utterance (Diessel & Coventry; 2021; Peeters et al., 2021). While a range of factors influences demonstrative choices across languages, in this thesis we are approaching the English demonstrative system and its multimodality while prioritising its basic, spatial function. The research in this line of argument, looking at this linguistic phenomenon from a spatial cognitive perspective argues a strong distinction between near and far space, within and outside reach of interlocutors (Coventry et al., in press) while keeping cross-linguistic differences in mind.

1.3 Deictic gestures

We now take a closer look at the nature of co-occurrence of gestures and demonstratives. From their early acquisition, demonstratives and pointing gestures are closely tied. Previous research suggests that the acquisition of language is preceded by the use of referential pointing (Levinson, 2018; Tomasello et al.,2005; Liszkowski et al., 2012).

1.3.1 Producing co-speech gestures

To better understand the multimodal nature of multimodal deictic communication, one can consider the nature of gesture and language binding more broadly. A large body of research has investigated the nature of gesture accompanying speech, debating its origins and processes involved in the utterances consisting of both words and gestures. The gesture accompanying speech has been regarded as part of a single process, as speech and nonverbal behaviour or gestures appear together "as a manifestations of the same process of utterance" as they translate intentions to communication (Kendon 1997;2011; McNeill, 1985, pp. 208 but also Butterworth & Hadar, 1989).

Gestures, just like utterances, are argued to be a depiction and expression of inner thoughts and intentions. Iverson and Thelen (1999) propose that gestures are embodied expressions of thought with arm movements, as a physical "manifestation of the embodiment of thoughts" (pp.19) while speech is a vocal expression of thoughts and mental activities from an early age. Freedman (1970) suggests that gesture/gesticulation also has an articulatory value for the speaker where gestures are "kinesic experiences" (pp.110) of inner representations of thoughts.

One should consider this multimodal, multifaceted nature of human communication as a linked system of expressing “lexical material” where the “burden of information can be shifted from one part to another” (Levinson & Holler, 2014, pp.1). While the origin of modern human communication and prevalence of vocal tract is rather new in evolution and is specific to humans (compared to other animals such as birds and bats), other systems of communication, such as ritualized gestures, emerged earlier in evolution and are similar to other primates (although mutual gaze and gaze following is argued to be not found in other animals). In fact, for the pointing gesture which this thesis is focused on (at least in English), the ritualization of the reaching movement has been argued to be the origin of pointing gestures in humans (imperative, declarative pointing), starting from infancy (Arbib, 2008, also see Frohlich et al., 2019).

Looking at the use of co-speech gesture and language in communication, Levinson and Holler (2014) argued for a layered system of communication with gesture capacity at the deepest level, later enriched with vocalisation and followed by a manual sign system (similar to gestural protolanguage) which then evolved to modern speech. Through this evolutionary view of the development of modern speech, we can also see how (from the beginning of language use as we know it now) communication had been multimodal. Similarly, Frohlich et al. (2019) argued for an evolutionary shift in the modality used for communication from gestures towards human communication today. Human communication showed similarities to the great ape communication, which evolved in time towards proto-language and to human communication today. The gestural and vocal streams refined over time and stayed consistent once took the form of recent human communication (Frohlich et al., 2019).

1.3.2 Development of referential acts

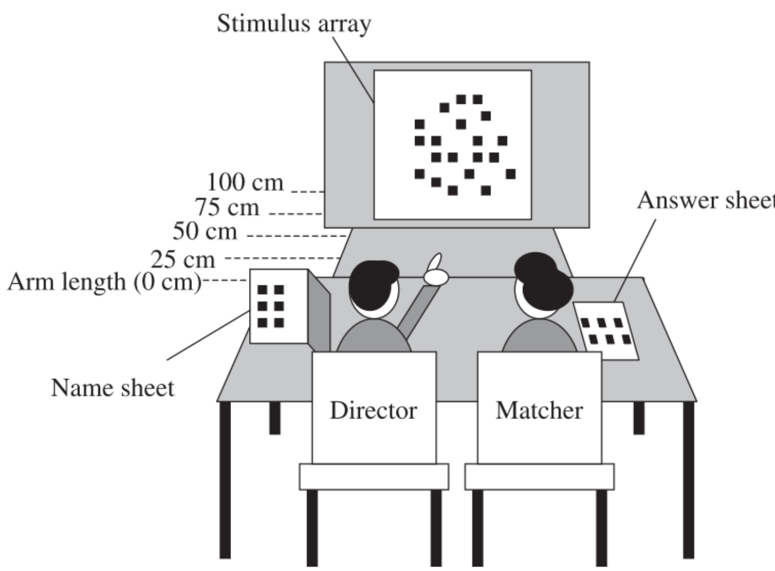
From an early age, infants engage in referential communication. According to Carpenter et al. (1998), infants first start to share and establish joint attention around 9 months which gradually develops until 15 months through following gaze and gesture. Furthermore, early development of joint attention (between 9-11 months) predicts the early emergence of gesture use to direct attention. Infants start following and comprehending intentions of gestures from 8 months in declarative form (Gredebäck, Melinder & Daum, 2010) and use the pointing gesture to direct and share attention and intention as early as 12 to 18 months (Liszkowski et al. 2004;2006), although previously Descrochers et al (1995) argued that children do not use pointing gestures intentionally until later in development.

Even after infants start to comprehend language, pointing tends to persist and be the preferential mode of communication in early development. Grassman and Tomasello (2011) argue that, when pointing and verbal information conflict, 2-to-4-year-old children prefer pointing over words when interpreting referential acts (also see Gonzalez-Pena, 2020). However, the types of gestures used by children and infants and their function change over time (similar to the development of spoken language from babbling to the first words used with meaning) starting from rhythmic, repetitive movements towards communicative/intentional gesture use at around 10-11 months (Iverson & Thelen, 1999). Nevertheless, Liszkowski et al (2011) argued that prelinguistic pointing starts between 10-14 months universally, and in accordance with Carpenter et al. (1998), interactions with the caregiver (and age of the infant) were implied as predictors of gesture use across different culture groups (ethnicity and languages), supporting the universality of prelinguistic gesturing. According to Diessel (2013), deictic pointing is then followed by demonstrative use as demonstratives are one of the first words children utter.

Considering the co-occurrence of pointing and demonstrative use from an early age, it has been proposed that the use of deictic pointing is part of the lexicalisation of demonstratives. According to Terengi (2022), "deictic co-speech gestures" are "spell-out part(s) of the internal structure of exophoric demonstratives: concretely, they contribute the spatial coordinates… that identify the location of the demonstrative's referent and deictic centre" (pp. 225) thus emphasizing the fundamentally multimodal nature of deictic expression. Deictic co-speech gestures are therefore a physical part of demonstrative forms, fully integrated into the syntax of demonstratives in a multimodal fashion. Terenghi (2020) further suggests that pointing gestures are the "vectorial component of demonstrative form" (pp.231). Such an argument is consistent with the claim that pointing obligatorily accompanies specific demonstrative forms in some languages (see Diessel & Coventry, 2020 for a review).

In an experiment looking at the nature of the co-occurrence of pointing and demonstratives, Bangerter (2004) engaged participants in a referential task. In this study, participants cooperated to identify target faces in an array of stimuli using gestures and verbal descriptions. The position of the array varied in distance, from a position within reach to 100 cm (Figure 1.7) from the interlocutors. There were two participants in each session, one would have a list of target faces with their associated names and guide the other participant towards the target faces (director) and the other participant would try to understand the face the director was referring to (matcher), then on an answer sheet, the write the names next to the faces. Bangerter (2004) also manipulated the interlocutor visibility as a condition where participants would not see each other, thus not have gestural information, but could hear each other as they verbally described the location of the faces. The gestures used by participants were visible to the experimenters. Behaviours of the director and matcher were recorded and coded for demonstrative pronouns and adverbs (*this*, *that*, *here*, *there*). From the set of gestures participants used, a gesture was coded as pointing as long as the referring arm was stretched out.

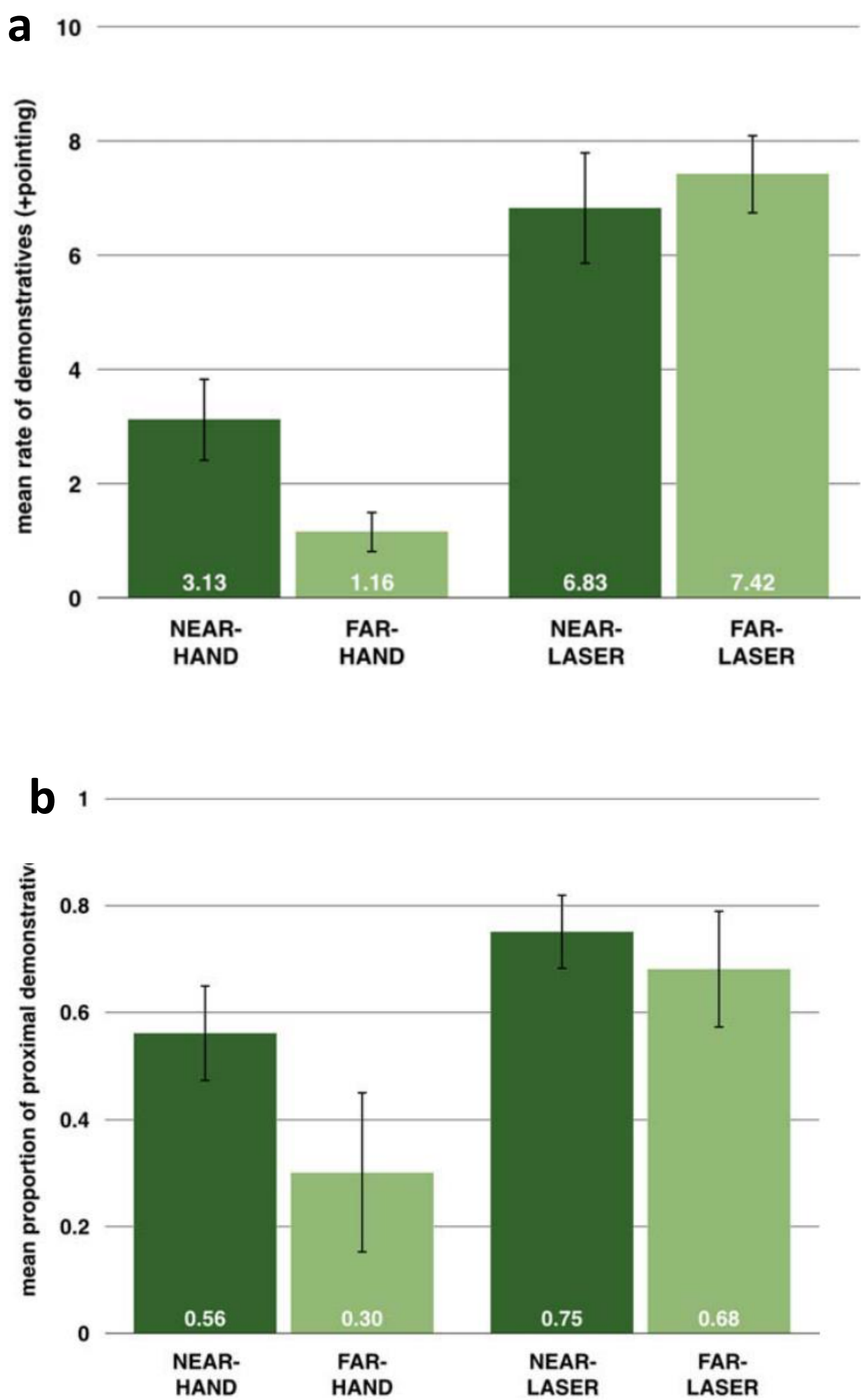
Figure 1.7: Experimental set-up of Bangerter (2004).
Participants were assigned the role of director or matcher. The director had a sheet with faces and their associated names, and the matcher had an answer sheet with only faces but no names. Participants saw the stimulus array with faces, some of which were the target. The stimulus array varied in its position relative to the participants.



The results suggested that when the stimuli array was closer, participants were more likely to use a combination of demonstrative and pointing gesture (although use of pointing overall did not vary as a function of distance) compared to the condition when the array was further away. Moreover, when participants could see each other, therefore see the pointing gestures used, they were less likely to use words when the target was closer (within arm's length) while pairs who did not see each other did not vary in the number of words they used as a function of distance. Therefore, Bangerter (2004) argued that, since in conditions with visible gestures the use of words varied as a function of increasing distance, the use of a pointing gesture led to reduced verbal effort (also see de Ruiter, Bangerter & Dings, 2012; Louwerse & Bangerter, 2005; Peeters, Azar & Özyürek, 2014).

Cooperrider (2016) built on the findings of Bangerter (2005) and further integrated the ambiguity of pointing as a factor. In this study, similar to Bangerter (2005), one participant would refer to the target in the stimuli array (which was either near or far from the participants) with their associated name and the second participant would try to associate the names with the correct targets (creature-like objects with different colours). The position of the stimulus array also varied. Participants would either point with their arms (ambiguous pointing) or with a laser pointer. A higher combined use of demonstratives with pointing arm was found when referring to creatures (both near and farther from hand) compared to the laser pointer. Furthermore, the participants used proximal demonstratives more often when pointing and distal demonstratives more often when not pointing. Supporting the mapping of space to demonstratives, Cooperrider (2016) also found that both in near hand and near laser conditions participants were more likely to use proximal demonstratives compared to distal hand and laser conditions although this effect was stronger in laser pointing (the less ambiguous option of the two) (Figure 1.8).

Figure 1.8: Results from Cooperrider (2016).

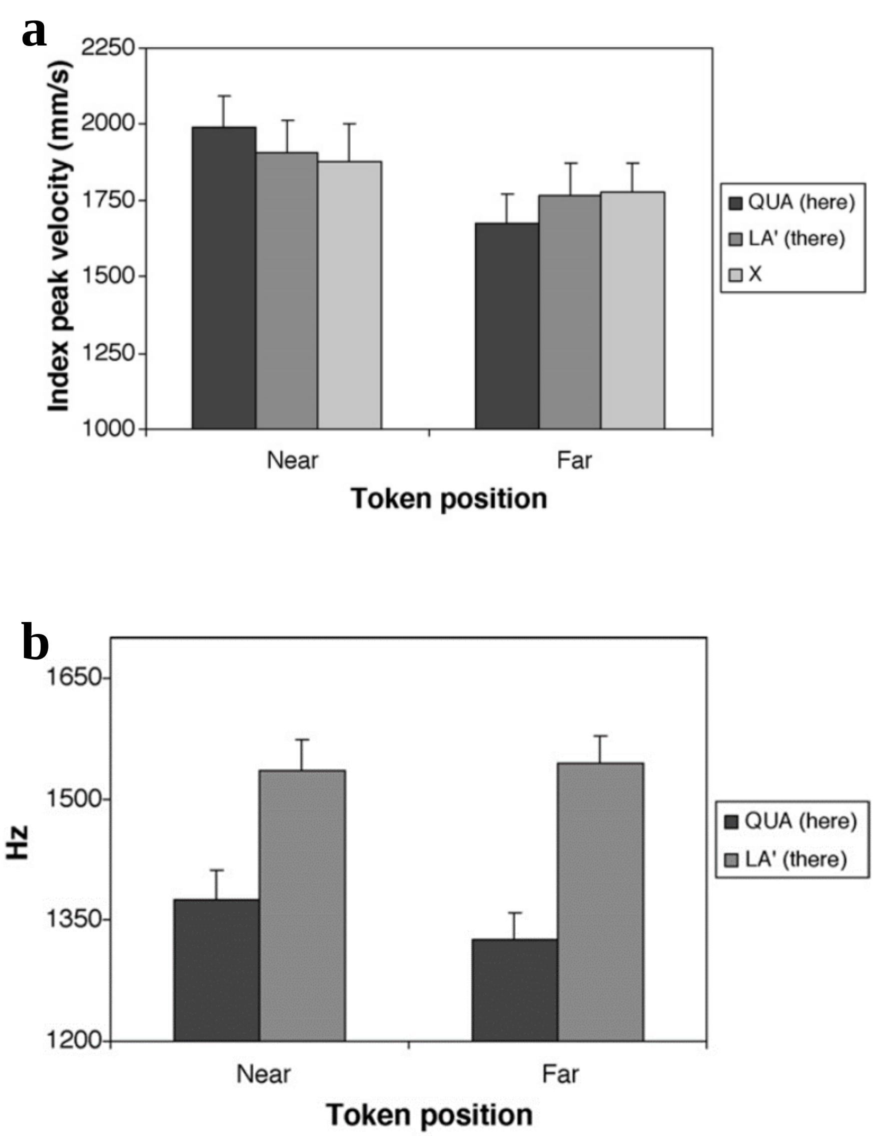


We now have established that gesture and demonstratives often co-occur in referential expression. Their use is influenced by environmental as well as psychological factors. Research also suggests that these factors do not only influence whether gestures and demonstratives co-occur but also the kinematics and acoustics of gestures and utterances.

For example, Chieffi, Secchi and Gentilucci (2009) manipulated the targets of demonstratives and pointing gesture to see the effect of incongruency on the physiological property of these deictic expressions. In two experiments, they asked participants to read out a deictic utterance on cue cards (termed "tokens" by the authors) and to gesture based on the position of the cue cards (near or far). The tokens had *here*/*there* written on them in Italian ("Qua" and "La"). When the token was placed near the participants, they would point towards themselves and when the token was farther away, participants would point forwards (termed "remote position" by the authors). As a control condition, participants would gesture (towards themselves or forward based on the position of the token) without any utterance when the token had X written on it. The demonstrative form written on the tokens and the position of the tokens varied in their congruency, for example, the token would be near the participant yet have the word *there* written on it, so the participant would gesture towards themselves while uttering a phrase indicating a farther position. The authors measured both the gesture kinematics and the voice spectra across six conditions. Pointing was measured via motion tracking with markers placed on participants' index fingers. Voice spectra were measured by computing a spectrogram of deictic words and further analysis was performed on the (average) time course of formant 1 (F1) and formant 2 (F2) of demonstrative words. Results suggested that when gesture and utterance were congruent, gesturing was faster compared to incongruent conditions and slower in incongruent conditions compared to control trials (Figure 1.9). Similar to the congruency effect on gestures, incongruent trials influenced voice spectra where F2 decreased when utterance and gesture were incongruent.

Figure 1.9: Results of Chieffi et al. (2009).

a) Experiment 1 gesture velocity. b) Experiment 1 voice spectra.

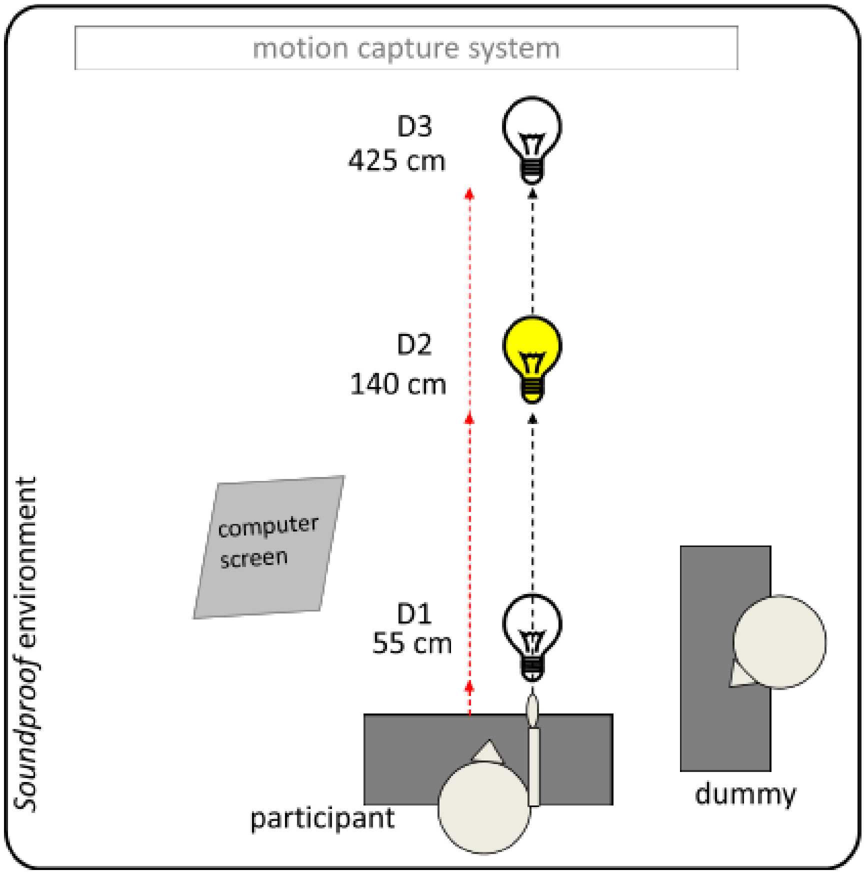


Similarly, Gonseth, Vilain and Vilain (2013) studied the gesture kinematics and articulation of verbal deictic expressions (manual pointing with the use of gestures and vocal pointing with linguistic elements of deixis such as demonstratives) in two experiments. They argue that the verbal and the gestural streams of communication cooperate and influence how each was expressed. The authors asked participants to use either gesture, speech (deictic reference in French) or a combination of the two to refer to a target. The targets were flashing lights located at three distances: within peripersonal space, 140 cm from the participant, 425 cm from the participant (see Figure 1.10 for the set up). Next to the participants, there was a "listener" which was a dummy in Experiment 1 and a confederate in Experiment 2. Gesture was measured in terms of kinematic properties (e.g., velocity) while the measure of speech and articulatory properties included lip opening and acoustic data of deictic utterances (fundamental frequency (F0) and formant values (F1 and F2 in Hertz).

Results suggested an effect of target position on pointing kinematics where pointing to a distal target was longer, faster and had a high amplitude. Unimodally, gesture hold was longer compared to conditions where participants could use both speech and gesture, and in this bimodal condition, the gesture use was faster. There was no difference in the gesture amplitude of the two conditions. The effect of distance was the same for the articulatory properties of deictic utterance with variations depending on distance, where lip opening was larger for distant targets. Lip opening was also larger in unimodal conditions. Regarding acoustic measures, F1increased as a measure of distance in Experiment 1. There was a unimodal-multimodal difference in utterance acoustics as a function of the availability of deictic expressions only in Experiment 2 (Also see Kita & Özyürek, 2003).

Figure 1.10: Set up of Gonseth et al. (2013) Experiment 1.

The participant saw flashes of light in three varying positions (D1, D2 and D3). Participants were cued to act based on the indications on the computer screen.



To sum up, the gesture and demonstratives are closely tied from acquisition to use in later life. They do not only co-occur, but their availability and congruency will also influence each other's expressions both in terms of likelihood of (co-)occurrence and in terms of how each are executed physiologically and acoustically. In what follows, we will have a look at the second non-verbal deictic expression, eye gaze, and its role in deictic communication.

1.4 Role of gaze in (deictic) communication

Eye gaze is a deictic expression used for directing attention to a specific position in space from as early as 12 months old, (Hanna & Brennan, 2007; Hanna, Brennan & Savietta, 2020; Thoermer & Sodian, 2001). It is also integral to ensure that all the interlocutors are engaged in the dyad, contributing to dyad-related processes such as turn-taking (Sacks, Schegloff & Jefferson, 1974; Vertegaal et al., 2001) and establishing joint attention (Carpenter et al., 1998; Diessel, 2006). It has been argued that eye gaze has spatial and social cognitive functions in deictic communication (Kendon, 1967; Kendon & Cook, 1969; Rayson et al., 2019; Stukenbrock, 2020).

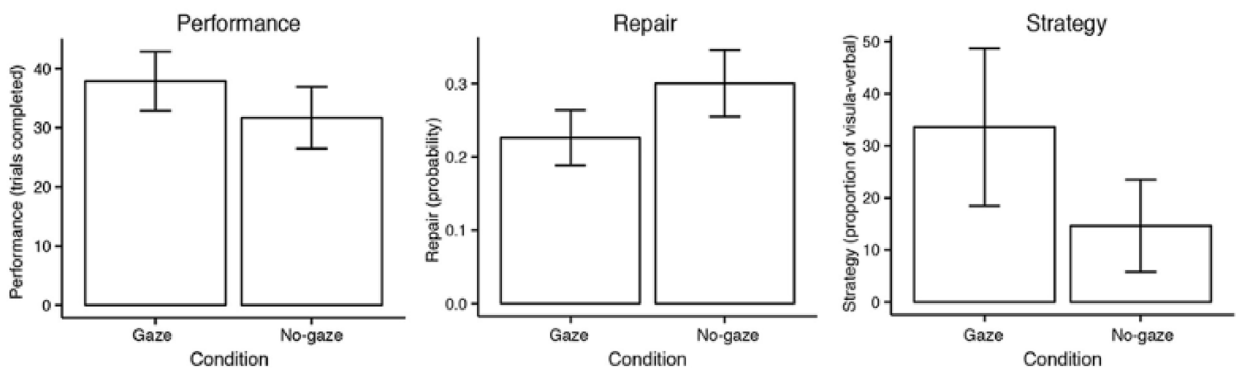
Streeck (1993) argues that gaze, gestures and language cooperate in communication. During gesticulation, gaze is exchanged between interlocutors, and during discourse, the target shifts between the interlocutor and the object targets. Gaze and gesture not only cooperate but gaze can replace pointing, "gesture and gaze are utilized in two fashions that, although they are interrelated, Cita [speaker] employs her hands to make pointing gestures, at other times, the function of pointing is entirely taken over by her eyes while her gestures become the objects of attention" (p.286). Holler et al. (2014;2015) further suggested that gaze can also modulate speech comprehension (also see Macdonald & Tatler, 2013).

Looking at the role of eye gaze specifically in referential communication, Perea Garcia, Ehlers and Tylen (2016) manipulated the availability of eye gaze information during a cooperative, spatial communication task. In this experiment (somewhat similar to Bangerter and Cooperrider's work on co-speech deictic gesture), one participant (director) would guide the other (matcher) to place objects (cups) in their designated positions. The director had access to a map with object position configurations and instructed the matcher to put the cups in these positions. A trial of object positioning was completed once all the cups were placed where they belonged, and the roles of matcher and director were swapped every four minutes. The director was not allowed to use any gestures or hand movements. Crucially, the availability of gaze was manipulated, and in the experimental condition participants wore goggles which allowed them only to see the stimuli but not each other's gaze, while participants in the control condition had full visual access to stimuli and eye gaze.

Participants' behaviours were video recorded for analysis. In the analysis, coders looked at task performance (number of trials completed within the time limit), verbal repair ("any expression that subsequently gave rise to instances of repetition of the same instruction, clarification or reformulation of the instructions using a different strategy") and strategy used (use of verbal, visual or a combination of these modalities) when referring to cups as they were placed to their designated position by the matcher. Results suggested that (Figure 1.11) performance without gaze was worse compared to the control condition, with higher use of verbal repair, therefore having to use speech as a modality more in the absence of gaze. Regarding strategy used, when gaze information was available, participants used a combination of verbal and visual information more often compared to the control condition. Therefore, the absence of gaze decreased the efficiency of communication with increased ambiguity in the descriptions and an increased reliance on verbal description.

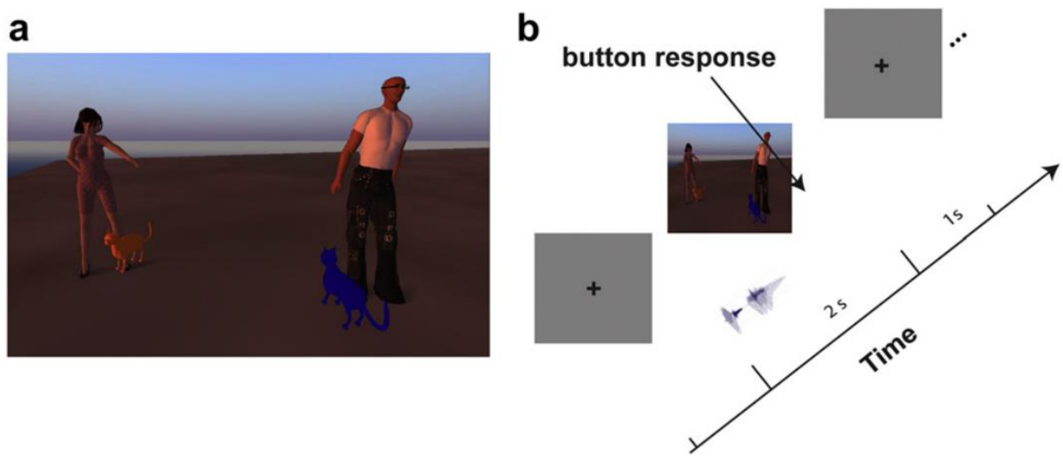
Figure 1.11: Results of Perea García et al. (2017).

Bar graphs representing performance measured by the number of trials completed, repair strategy in terms of likelihood of using verbal repair and strategy used by participant measured in likelihood of using a combination of visual indication and verbal description across two conditions: with and without eye gaze information.



The importance of eye gaze in deictic communication (and joint attention) was also supported by Stevens and Zhang (2012) in an electrophysiological and behavioural study (Figure 1.12). In this study, authors manipulated the semantic congruency of visual scenes and auditorily presented demonstratives. The images contained information on the position of referent and interlocutors, (referent either close to the speaker, listener or far from both), and whether they established joint attention via gaze (shared gaze or not while the speaker was always pointing at the intended referent). For the behavioural measure, participants were asked to push buttons to react to congruency. Congruency of the stimuli was achieved by manipulating the target of speech and gaze. The image seen and the deictic utterance were congruent when, for example, near-gaze was coupled with "*this* one" and when incongruent, near-gaze was coupled with "*that* one". The EEG data were collected with 64-channel system for ERP analysis.

Figure 1.12: Experimental paradigm used in Stevens and Zhang (2012)
a) the visual stimuli example. The images had two potential referents (orange cat and blue cat) and the speaker would always point towards the intended referent. b) The paradigm of image and auditory stimuli ("this one" or "that one") coupling. The scene coupled with simultaneous sound lasted two seconds.



The behavioural results suggested significant effects of referent position and congruency of gaze and utterance and a significant interaction between referent position and shared gaze. There was an effect of congruency only in the conditions where the interlocutors shared gaze, and when the referent was near the speaker with slower response times for incongruent pairings compared to congruent pairings. When the gaze was not shared, there was no effect of congruency. Electrophysiologically, as well as joint attention established by gaze, the proximity of the referent to the hearer played an important factor while proximity to the speaker did not matter. The ERP results suggested a significant effect in the 525-572 window for referent position, shared gaze and laterality. Additionally, there was a significant interaction between referent position and modality congruency. This interaction implied an effect of shared gaze when the referent was placed near the hearer but not when near the speaker or far from both interlocutors which is not particularly in line with behavioural data. These results overall emphasized the importance of joint attention and the position of the hearer in deictic reference. (see Stevens and Zhang (2014) for gestural adaptation of the study).

1.4.1 Interim summary

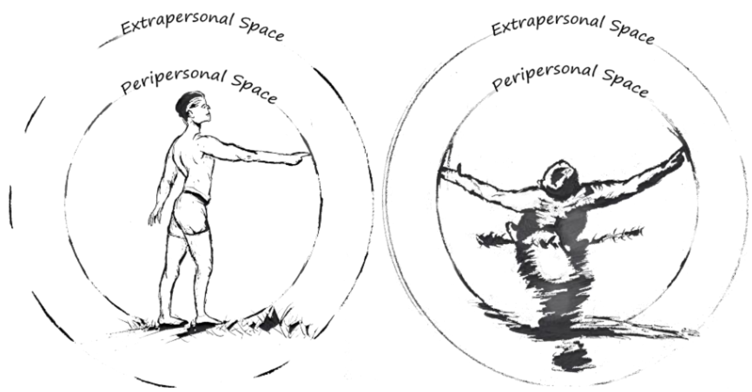
In deictic communication, spatial demonstratives, gestures and eye gaze co-occur to indicate a specific position in space. Processing of communicative intent can be somewhat more efficient when all the deictic expressions are available to interlocutors simultaneously. The way deictic expressions are used is influenced by an interplay of environmental, physical, psychological and cognitive factors. One particular factor which consistently influences demonstrative choices in all languages is referent proximity.

In what comes, we move from consideration of the multimodal nature of *communication* to consideration of the multimodal nature of *perception* underlying spatial communication. We will have a brief look at the literature related to the secondary interest of this PhD; the implications of the multisensory nature of peripersonal space on demonstrative choices. As discussed above, perceived closeness is a strong determinant of demonstrative choices. While the perception of space is multisensory, most of the research on the mapping of demonstratives to space has focused on the visual perception of space, without considering the possible implications of use of non-visual senses, such as haptics, on demonstrative choices. Therefore, we are yet to understand whether there are any implications of the multisensory nature of peripersonal space (Figure 1.13) on deictic communication (further elaborated in Chapter 3 and 4).

1.5 Multisensory implications of deictic communication

The world within an arm's reach (peripersonal space) can give us access to an immense amount of (sensory) information at a given point in time (Noel et al., 2018; Pellegrino & Làdavas, 2015; Serino, 2019; Salomon et al., 2017; Wallace et al., 2004). In harmony, sound, sight, smell, and touch inform us about object properties and positions in space (van Beers, Sittig & Denier van der Gon., 1996; Driver & Spence, 2000; Huo et al., 2023; Neil et al., 2006; Schifferstein, Smeets & Postma., 2010; Serino, 2019).

Figure 1.13: Depiction of peripersonal and extrapersonal space. *Courtesy of S. Umurtag-Kocabiyik.*

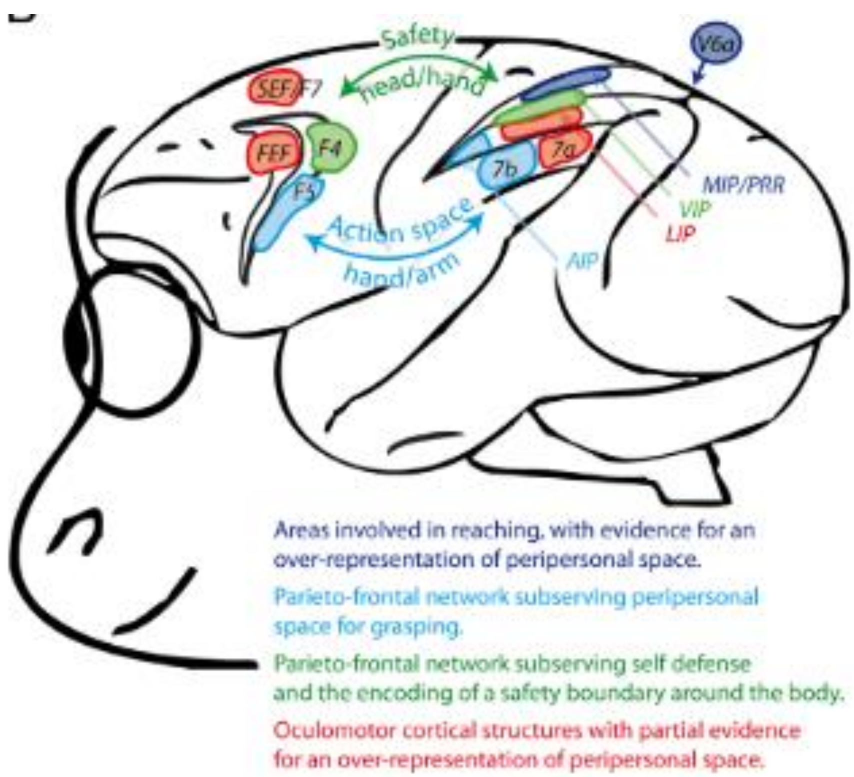


Our brains integrate information from a range of senses to create a coherent representation of space (Deneve & Pouget, 2004; Gori et al., 2022; van der Stoep, Postma and Nijboer, 2017). Different types of sensory input are processed from varying reference frames at a given time. However, we do not have access to this information individually and our "experience of sense is not fragmented" (Serino, 2019: p138), instead, senses are integrated unconsciously to create a unified spatial representation (Salomon et al., 2017). Holistically, spatial information from different senses becomes a part of and a tool for higher cognitive functions (Gori et al.2022; Morimoto, 2022; Neil et al., 2006; Serino, 2019) and influence motion (Brozzoli, Ehrsson & Farne, 2013; Macaluso & Maravita, 2010), action (Barutchu, Crewther & Crewther, 2009; Finisguerra et al., 2014; Làdavas & Serino, 2008; Maravita, Spence & Driver, 2003; Serino et al., 2007), and communication (Arbib, 2008; Coventry et al., 2008; 2014; Rizzolatti & Arbib, 1998).

One sense that has been primarily used to understand the role of spatial perception in demonstrative choice is vision, which is also argued to be the most dominant sense for spatial perception. The sense of vision acts as a calibration for other, less reliable senses for spatial perception such as proprioception (reaching to grasp objects), touch or sound (Klatzky & Lederman,1993; 2003; 2011; Kolarik et al., 2016; 2020; Pasqualotto, Finucane & Newell, 2005). In the absence of vision, there are distortions in the way a target location is perceived (see Chapter 4 for more information).

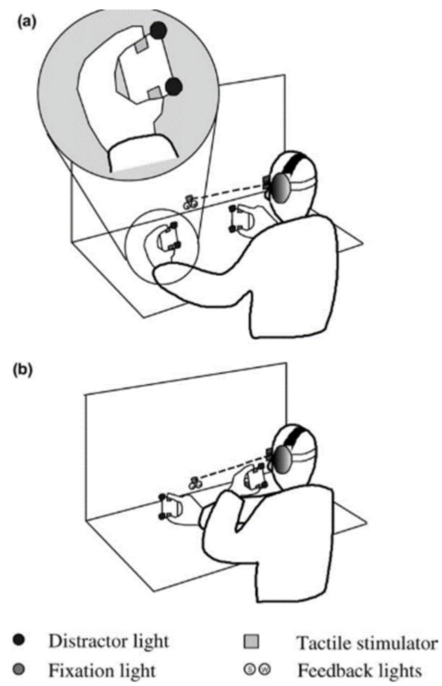
Experiments supporting the multisensory nature of peripersonal space come from a range of quarters, including primate models, behavioural and neuroscientific studies with neurological patients (e.g., crossmodal extinction patients (di Pellegrino et al., 1997)) and healthy populations (di Pellegrino & Làdavas, 2015; Cléry et al., 2015; Serino, 2019). Neurological studies suggest the involvement of parietal, premotor and prefrontal periarcuate regions (Cléry et al., 2015) and the putamen (Macaluso & Maravita, 2010; Serino, 2019) in the processing of stimuli from multisensory channels in peripersonal space. Primate models (e.g., single neuron recordings, fMRI studies etc.) suggest that a collective group of neurons respond to and integrate information from different senses simultaneously within peripersonal space but no other position away from the body. When there are multiple senses available from the same location, perception is enhanced (Macaluso & Maravita, 2009). Based on collated evidence from research concerning peripersonal space, a general role of the dorsal network for near space (left dorsal occipital cortex, left intraparietal cortex and left ventral premotor cortex) and the ventral network for far space (bilateral ventral occipital cortex and right medial temporal cortex) is implied (Cléry et al., 2015 but also see Weiss et al., 2003). Overall, Cléry et al. (2015) propose that "arm and hand related actions shape cortical peripersonal representation", (p.140) indicating how areas relating to peripersonal space also relate to manipulating objects via e.g., grasping (see Figure 1.14, for representation of these areas) and one might have the urge to assume, even gesturing (also see Brozzoli et al., 2011).

Figure 1.14: Representation of functional networks involved in peripersonal space, from Cléry et al., (2015).



The behavioural studies looking at the multisensory nature of peripersonal space come from the experiments modulating availability, congruency and position of stimuli across different senses. For example, Spence et al. (2000a, b; 2004) describe a "crossmodal congruency task" where the position of tactile stimulation and visual distractors were manipulated (Figure 1.15). In these experiments, the tactile target and visual distractor could either be at the same or the opposite locations. The (vibro)tactile stimulation would either be on the index finger or the thumb of either one of the participant's hands, while the distractors would be in the opposite or the same position as the tactile stimulation. The participants would indicate where the stimulation was, either their index finger or thumb (known as *elevation discrimination response* where the index finger symbolizes the above position the while thumb below).

Figure 1.15: Schematic representation of crossmodal congruency in Spence et al., (2004).



In studies using this methodology, typically, when vibrotactile and visual stimulation were incongruent, the accuracy of location judgement was lower, known as the crossmodal congruency effect (congruency effect = performance on incongruent distractor trials minus performance on congruent distractor trials). Based on the overall results from studies using this methodology, the authors argue that, while vision is a distractor for stimuli localisation when the two senses are incongruent, it can be a facilitator when the vision and tactile stimulation are congruent. Similar findings of a congruency effect are also found in tactile discrimination tasks with auditory distractors (Serino et al., 2007; 2011; Spence, Randon & Driver, 2000. Also see Serino et al., 2017 for detailed discussion of multimedia measurements for multisensory space). The effect of congruency in visuotactile stimulation is also supported by electrophysiological studies where a late positivity at the posterior parietal cortex for congruent visuotactile stimulation (of/around left hand only) was observed (Longo, Musil & Haggard, 2012). Based on findings with behavioural tasks as well as primate and human neurological findings, we can see that in the peripersonal space, information across different senses influences each other, and when congruent and available, enhances spatial perception.

Another interesting characteristic of peripersonal space is its plastic nature. The distinction between peripersonal-extrapersonal space is not a categorical near versus far distinction, but rather there is a graded fall-off from peripersonal space to extrapersonal space (Bufacchi & Ianetti, 2018). The boundaries of peripersonal space can be extended with (and without) tool use (Làdavas & Serino, 2008; Longo & Lourenco, 2006; Magosso et al., 2010; Serino et al.,2007;2015) and even the simple act of grasping the target was argued to remap peripersonal space, almost similar to the effects of tool use (Brozzoli et al., 2009). Previously, Coventry et al. (2008) in fact found an effect of tool use on use of spatial demonstratives where the word *this* was used more often, even for positions outside reach of the arm, when participants pointed at objects using a tool (mirroring effects of tool use and the extension of space: see Berti & Frasinetti, 2000).

1.5.1 Implications of the multisensory nature of space on demonstrative use

As mentioned earlier, almost all research to date looking at the interplay of spatial perception and demonstrative choices has been based on vision. If indeed peripersonal space or spatial perception overall is multisensory, and if the sense of vision is the dominant sense, acting as the calibrator of any other sense, we would expect changes in perception in the absence of vision. Therefore, demonstrative choices in the absence of vision might be different compared to situations in which spatial resources are available from multiple channels including vision. In Chapters 3 and 4, I will come back to the multisensory perspective on deictic communication, looking at a population that relies on non-visual spatial experiences in their daily lives - blind and visually impaired individuals.

Furthermore, if peripersonal space boundaries are indeed plastic, which can be extended both with tool use and action, what would the implications of unimodal-bimodal perception of space be for the use of demonstratives? For example, would reaching forwards to feel an object's position increase the peripersonal space, and lead to an increase in the use of proximal demonstratives?

1.6 Precis of the Thesis

Building on the current knowledge of deictic communication regarding its multimodal nature and multisensory characteristics of peripersonal space as reviewed above, the aims of this PhD were twofold. First, in four experiments, we investigate the multimodal nature of deictic communication, closely examining the relationship between demonstratives, gesture, and eye gaze in referential communication. As we briefly reviewed earlier, some studies have presented evidence for a close interplay between modalities in deictic communication, but, to our knowledge, no study to date has looked at the relative role of pointing, eye gaze and linguistic cues in directing attention to a spatial referent. We probe for the relative importance of deictic expressions by manipulating their availability and congruency. Finally, we also manipulated the stimuli type where participants saw an individual engaging in the referential act both as images coupled with audio (of demonstrative use) and videos of the utterance. I would like to remark that Experiment 2 was a pilot for Experiments 3 and 4. Therefore, here we only elaborate on the methodology and results of Experiment 2. Experiments 1, 3 and 4 are, currently under review in peer-reviewed journals (detailed in Author Declaration) and are presented in manuscript form accordingly.

The second focus of this thesis considers the implications of the multisensory nature of spatial perception to the use of spatial demonstratives. As reviewed above, perception of space is an important determinant of spatial demonstrative use, where referents within reach are labelled with proximal demonstratives more often. Given that research on peripersonal space suggests that spatial perception within this region is very rich, consisting information from almost all senses simultaneously, we deliberate on sensory implications for the spatial perception-demonstrative mapping. For this purpose, we developed a haptic adaptation of the memory game paradigm (earlier versions described above) to be used in two experiments probing the availability of visual and haptic sensory information. In Experiment 5, sighted participants complete the task using haptic and/or visual information to perceive the referent position. Experiment 6 is a pilot study using the same methodology with a blind and visually impaired population. In the final chapter, I will review the findings across six experiments in the broader literature, with implications for future research.

.

Chapter 2 Multimodal Characteristics of Deictic Communication

In what follows, I will present four experiments looking whether there is an advantage of seeing multimodal deictic cues in guiding attention or if language alone is a sufficient cue for reference. We will also observe whether one deictic expression is stronger than the other. In these experiments, people saw the same images. In the first experiment, we manipulate gesture, eye gaze and linguistic reference (a noun referring to one of the objects) to observe the effects of these variables on the demonstrative choices of the participants. In rest of the experiments, we manipulate gesture, eye gaze and demonstratives as cues to reference. The levels of non-verbal cues (pointing and gaze) and the items presented in the images were the same (as in Figure 2.3). Experiment 2 served as a pilot for the experiments 3 and 4, and therefore only the methodology and results are reported (without discussion of the findings).

Please note that the following sections are taken from the papers submitted for publication (as described in the Author's declaration). The figure numbers and order of figures have been changed for clarity and to improve the flow of experiments. Please note that the introductions serve as a recap of literature reviewed earlier as a basis for topics discussed.

2.1 Experiment 1

2.1.1 Introduction

Spatial demonstratives are among the most used and oldest words in all languages (Diessel, 2006; Diessel & Coventry, 2020). The acquisition of demonstratives is also an important milestone in language development since they are among the first words children learn (Clark & Sengul, 1978; Diessel & Monakhov, 2022). The number of available demonstrative words, the ways demonstratives are used non-spatially (Rubio-Fernandez, 2022) as well as the shape and types of the accompanying gestures might vary across cultures (Diessel, 2006, 2014; Enfield, 2003), but the use of deictic reference to establish joint attention is universal.

From an early age, pointing, eye gaze and demonstratives co-occur synchronously and almost simultaneously to capture and direct attention to an object (referent) in space (Bangerter, 2004; Carpenter et al., 1998; Piwek et al., 2008; Todisco et al., 2020). For example, Todisco et al., (2020) observed caregivers and children interacting while reading a storybook. Analysis of the deictic episodes showed that interlocutors use speech, gesture, and eye gaze synchronously in deictic communication most of the time. As described in Chapter 1, Bangerter (2004) suggests that participants tend to use pointing with demonstratives when the referent is proximally located, with an associated overall reduction in verbal effort and word use for the objects within arm's reach (see also Louwerse & Bangerter, 2010). In contrast, the use of gesture was reduced as the distance of the target increased while the verbal description of the object's position and mention of features of the referent increased (Bangerter, 2004). Building on these findings, Cooperrider (2016) manipulated how participants pointed to examine potential changes in the use of pointing and demonstratives in combination. In this experiment, participants either pointed to the target with their arms (the ambiguous option) or using a laser pointer (the less ambiguous option). Consistent with Bangeter (2004), the results suggested that participants were more likely to use *this* than *that* when pointing and that more than this when not pointing. However, when pointing was ambiguous, the combined use of demonstrative and pointing was less frequent compared to non-ambiguous pointing.

For some languages, it has been claimed that pointing is obligatory when using demonstratives (see Diessel & Coventry, 2020 for a review). For example, in Yeli Dnye, demonstrative use requires either pointing and/or gaze (Levinson, 2018:32). Other languages with obligatory gestures include Yucatec (Bohnemeyer, 2018), Warao (Hermann, 2018) and Tiriyo (Meira, 2018). Terenghi (2022) argues that deictic gestures are fully integrated into the syntax of demonstratives as their "vectorial component", even for English, and therefore might not merely co-occur with demonstratives (exophoric form) but be a part of the internal processes involved in demonstrative use (also Kowadlo, Ye & Zukerman, 2010 and Wagner, Malisz & Kopp, 2014, for an overview of other speech accompanying gestures). The early co-occurrence of demonstratives and pointing also supports the lexicalisation of demonstratives with pointing gestures (Clark & Sengul, 1978; Diessel, 2006; Iverson & Goldin-Meadow, 2005; Liszkowski, 2010). From 9 to 12 months, infants start to follow gaze and use gestures and actions to draw the attention of others (Carpenter et al., 1998; Liszkowski et al., 2004; Tomasello 2003). Eye gaze is also argued to be an important aspect of deictic communication, particularly because listeners tend to use eye gaze as a visual cue to understand the point of reference (Hanna et al., 2019; Hanna & Brennan, 2007) and due to its social cognitive role in face-to-face communication (Diessel, 2014; Diessel & Coventry, 2020; Holler et al., 2014; Prasov & Chai, 2008) Furthermore, in some languages, use of gaze accompanying a demonstrative is obligatory (Diessel & Coventry, 2020). Manipulating the synchrony of eye gaze and demonstratives in an EEG experiment, Stevens and Zhang (2013) suggest that joint gaze (and the relative distance of the referent) is important in using spatial demonstratives. Perea García et al. (2017) further support the importance of shared eye gaze in referential communication. In their study, interlocutors did not have access to each other's eye gaze and the absence of eye gaze led to ambiguity and therefore difficulty in establishing a common referent (Please see Chapter 1 for a detailed explanation). Furthermore, even when participants resort to using more words to compensate for the absence of eye gaze, the establishment of a referent was not as efficient as in the conditions when the gaze information of the interlocuters was available (Perea García et al., 2017).

Leaving aside pointing and eye gaze, as I discussed in Chapter 1, there is a body of evidence to suggest that choice of demonstrative is affected by a range of spatial and non-spatial parameters (Diessel & Coventry, 2020; Peeters et al., 2020), including object distance (Anderson & Keenan,1985), location of the addressee (e.g. Jungbluth, 2003; Coventry et al., 2008; Rubio-Fernandez, 2022), the attentional status of speaker and addressee (Rubio-Fernandez, 2022), the positon of the object vis-à-vis enviroment (e,g, elevation; Forker, 2020) and properties of the object (e.g. visibility, ownership, etc.; see Coventry, Griffiths & Hamilton, 2014). Prominent among these is the location of the referent. Recently, it has been shown that reachability rather than mere distance affects the choice of *this* versus *that* in English (Caldano & Coventry, 2019; Coventry et al. 2008, 2014). Moreover, the basic reachable/non-reachable distinction has been shown to be a fundamental structuring tool for the demonstrative systems of 29 diverse languages (Coventry et al., under review), and is therefore a viable candidate as a semantic universal.

While choice of demonstratives is affected by reachability and other parameters (Coventry et al., 2014; Peeters & Özyürek, 2016) it is also recognised that the demonstrative *that* also has a non-deictic, definite function, (somewhat similar to the definite noun phrases like *the*) making it a demonstrative which is “more sought after”, with a non-spatial function since it can presuppose a unique, distinct referent (Ahn, 2017; Diessel & Coventry, 2020; Ionin et al., 2012; Kaplan 1977 but also Wolter 2006 and Roberts 2002). In other words, *that* in English can be regarded as a ‘neutral’ (Levinson, 2018) or ‘default’ (Diessel & Monakhov, 2022) demonstrative that can be used for any object location. This begs the question of what determines the use of the default demonstrative in a spatial context? Specifically, one can ask, given the ubiquity of pointing and eye gaze with demonstrative use, and especially for the proximal form, whether the absence of gesture and or eye gaze flouts the requirement for the use of the proximal term, leading to the use of the default term, *that*, in the language.

In order to understand to what extent knowing the location of the referent, pointing and eye gaze are necessary for demonstrative choice, we manipulated the congruence of these three cues. In an online experiment, participants saw images of a person (agent) pointing and gazing towards one of two objects on the table (the referents) and these images were coupled with a noun representing one of the two potential referents (e.g., mango, Figure 2.1). Participants were asked to pick the spatial demonstrative (*this* or *that*) they thought the agent would utter, based on the location of the referent (as denoted by the noun), pointing and eye gaze information (Figure 2.2 and 2.3). By changing the referent location and with varying combinations of pointing and eye gaze (Figure 2.3), we explicitly tested two hypotheses. First, we tested whether knowing object location alone is sufficient for predicting demonstrative use , or whether congruent pointing and eye gaze is necessary for the spatial use of demonstrative terms (especially the proximal demonstrative; Bangeter, 2004). Specifically, if non-verbal cues are part of the lexical representation for the proximal term (e.g., Terenghi, 2022), we predicted that the absence of pointing and/or gaze would lead to an increase in the choice of the default or neutral demonstrative, *that*, when the object is proximal to the speaker. Second, we asked if incongruence between object location, pointing and eye gaze would lead to a change (is there a cost to mismatch) in the proximal-distal distinction for spatial demonstrative choices. We hypothesized that having more than one cue will lead to a more decisive choice of demonstratives. We also hypothesized that when pointing and/or eye gaze do not match object location participants will pick the proximal demonstrative this less often (or the distal demonstrative, *that*, more often).

2.1.2 Methodology

Participants: Participants were students at the University of East Anglia, taking part for course credit (N=38, 28 female; M_{age} = 19.7, SD=1.54). Power analysis was carried out prior to recruitment using G*Power (version 3.1) (Faul, Erdfelder, Buchner & Lang, 2009). We followed sample size recommendations of Bryasbaert (2018) and Faul et al (2009). To detect a medium effect (f = 0.25) at a significance criterion of α = .05 for power of 80%, a sample size of 36 participants for this experiment was indicated. Brysbaert (2018) argues these measures to be reliable guide when setting up new research, in absence of similar studies reporting reliable and consistently effect size. Additionally, available limited yet research, although not directly using same methodology but manipulating similar factors such as relation of referent distance and demonstrative words (Coventry et al, 2014), gaze and gesture availability/congruency on language processing and use (Cooperrider, 2016; Holler et al., 2014; Kelly, Özyürek & Maris, 2010; Kelly et al., 2015; Perea García, 2017) reported medium and medium-to-large effects and differing but mostly strong correlations amongst these factors. This basis for the power analysis was used for the following three experiments reported in this chapter.

Non-native speakers of English and participants who completed the experiment via their mobile devices were excluded from further analysis.

Stimuli and Design: Stimuli consisted of images of the agent (male) sitting behind a table, pointing (only left arm), and gazing at one of the two items placed on the sagittal plane; one positioned proximally (close) and one distally (far) relative to the agent. These images were coupled with a noun labelling one of the two potential referents (as information on object location; Figure 2.1). The task was to complete a two-word sentence by picking the demonstrative (*this* or *that*) they thought the person would utter. The location of the choices (left or right) was counterbalanced. Participants had unlimited time to choose a demonstrative.

Figure 2.1: An example trial from the experiment.

(a) proximal item, (b) distal item and (c) selection of demonstrative as a function of object location denoted by a noun.

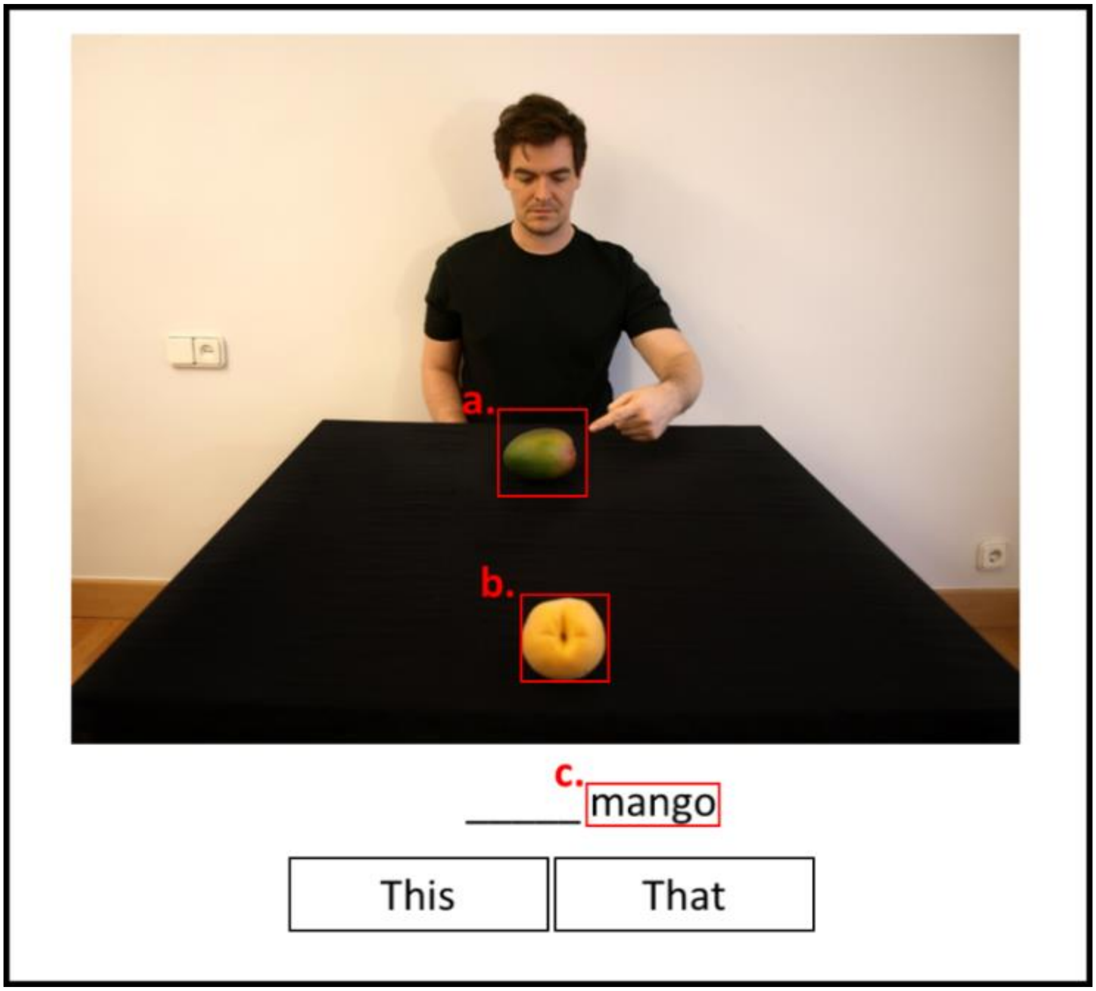
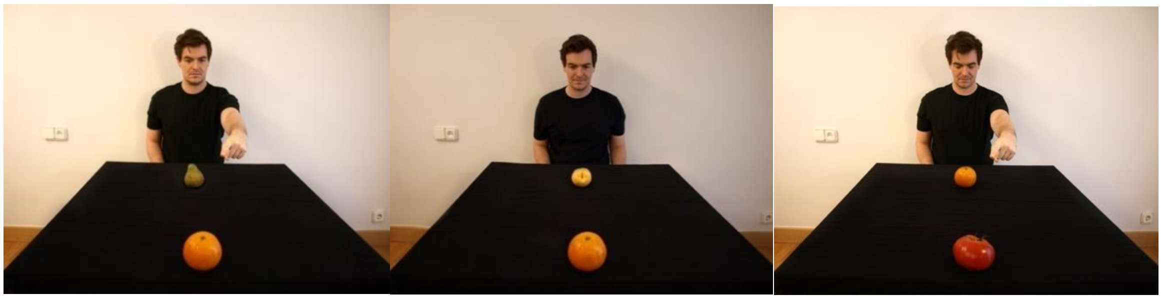
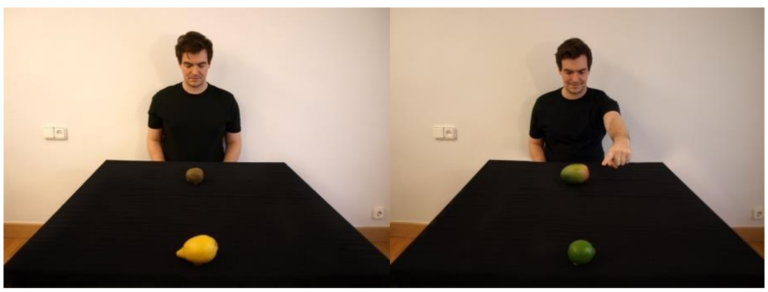


Figure 2.2: Examples of different conditions.

a) b) c)

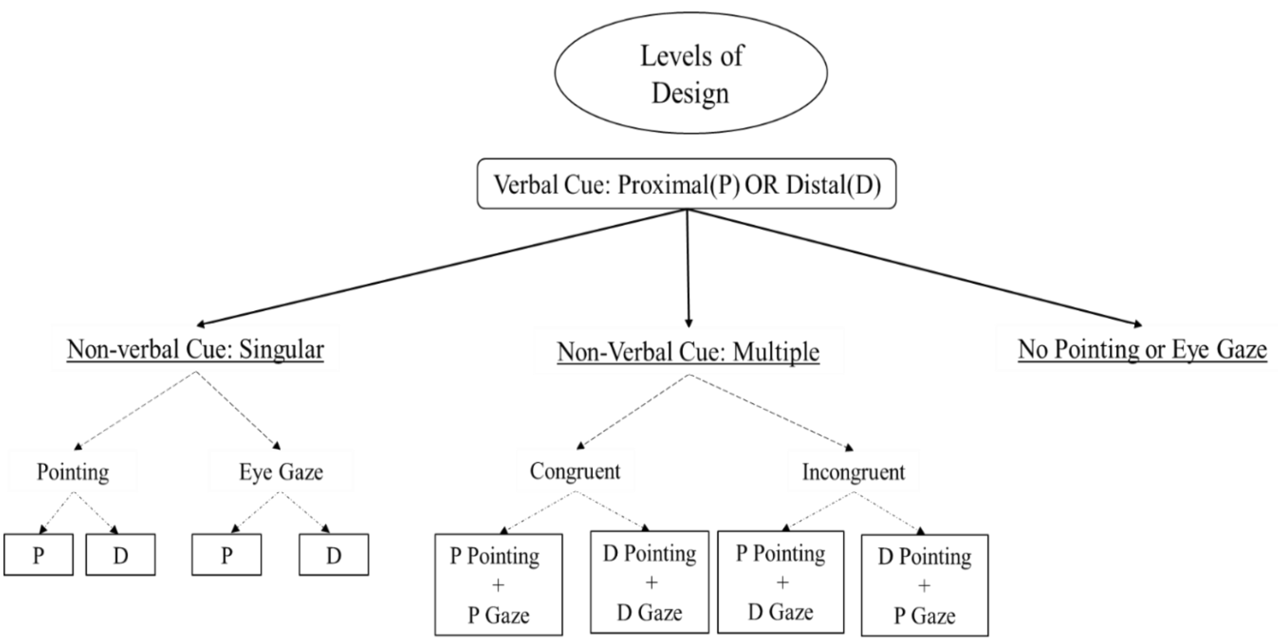


d) e)



a) Pointing + eye gaze congruent b) Eye gaze only c) Pointing only d) No pointing or gaze e) Incongruent eye gaze and pointing

Figure 2.3: Overview of the design. P=Proximal, D=Distal.



There were three manipulations in the experiment: object location, pointing arm and eye gaze. The noun would either represent the proximal or the distal item (Figure 2.3). For example in Figure 2.1, the noun is referring to the proximal item. The pointing arm varied across three levels; pointing at the item closer or farther relative to the agent and a third level when the agent was not pointing (arm resting on his leg). The third variable was the eye gaze of the agent, which (like pointing) targeted either the proximal or the distal item or was absent (the speaker's eyes were closed). The experiment employed a repeated measures design (2 x 3 x 3), with a total of 18 levels, consisting of a combination of each one of the operationalised levels of the three variables. While in some levels of the design (Figure 2.3), the pictured person did not point (no pointing) and/or close his eyes gaze (no gaze), there was always a noun to represent one of the two items. During the experiment, images in each trial were a unique combination of the objects, pointing and eye gaze and none of the images (with possible combinations) were repeated.

Each participant completed four experimental blocks, with a total of 72 experimental trials (with four data points for each level of the design per participant). The order of blocks and trials within each block were randomised between participants. After the four blocks were completed, participants were also given a gaze detection task (with a total of six trials per participant). The aim of the gaze detection task was to check how accurate participants were in detecting the gaze of the pictured person.

Procedure: The experiment was presented on the online experimentation platform Gorilla (Gorilla.sc). Participants were instructed that their task was to click on the word (a spatial demonstrative) to complete the two-word sentence they thought the person was uttering, based on the information available in the picture. After participants agreed to take part in the study, they completed an example trial. The example trial was a single experimental trial (the image was not used during the rest of the experiment) with written guidance on how to complete the task. The wording of the example trial was chosen carefully to avoid priming to any one of the cues. After the example, participants began the experimental trials, followed at the end by the gaze detection task. The task took 40 minutes on average. After the experiment, participants were fully debriefed.

Data preparation: Every time a participant used word *this* they were given a score of one (1) and zero (0) for *that*. The data were then analysed in terms of the likelihood of picking the word *this*.

Analysis Model: The experimental data were modelled and analysed using Generalized Mixed Models (Gallucci, 2019) in Jamovi (version 2.2; The jamovi project, 2021). Due to the binomial nature of the outcome (1 and 0), the model had a binominal distribution (of residuals) and logit link function. The outcome was computed as the likelihood of participants' picking the spatial demonstrative *this*. We included random coefficients with participants as intercepts, predictors, and predictor interactions as random slopes, as long as the model converged, and model fit was not singular. The random effect structure was formed considering the convergence and theoretical relevance (Barr, Levy, Scheepers & Tily, 2013). The final model included participants as intercept, spatial cue and pointing as random components. Fixed effect factors included pointing, eye gaze, spatial cue and all possible two-, three-way interactions. The model was estimated using log likelihood (odds ratio, OR) and OR is reported for model estimates. When the OR was less than 1, we report the inverse odds for interpretability (1/OR). The main effect and interactions of the predictors were reported from fixed effects omnibus tests. The fixed effect coefficients were assessed using Wald's Z statistics (Table 7.1). All factors were coded with Helmert contrast. The graphs and descriptive data reported are based on estimated marginal means. Error bars in the figures represent the standard error of estimated marginal means. Significant interactions were then further investigated with post hoc analysis with Bonferroni correction.

2.1.3 Results

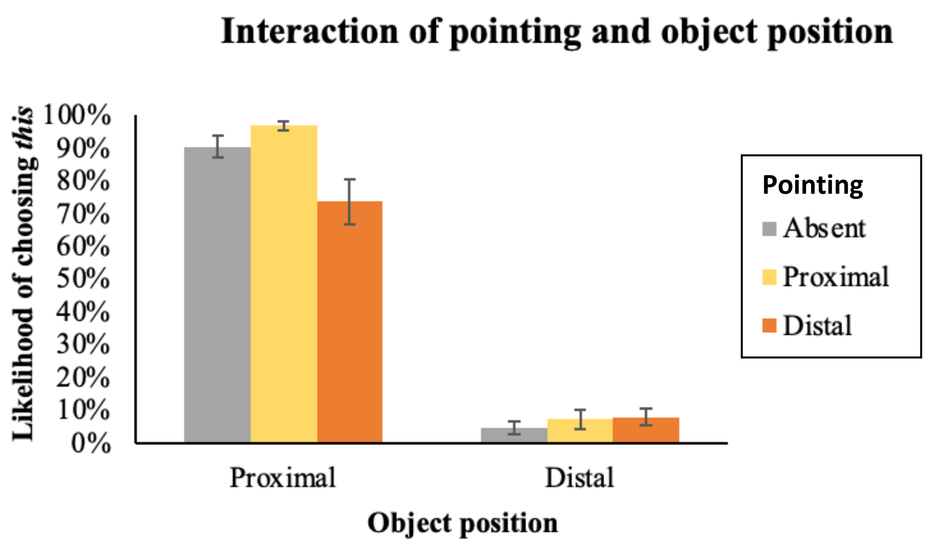
We first examined the gaze detection task to ensure participants were able to identify where the person in the image was looking. The highest score possible in the gaze detection task was 6. Average score for gaze detection accuracy was 5.74 (*SD* = .55) where 79% of participants detected gaze location in all the images. Therefore, none of the participants were excluded based on their performance in the gaze detection task.

The model testing the main effect and interactions (Conditional R^2= 0.76) of pointing, eye gaze and object location on spatial demonstrative choice produced significant main effects of object location (noun), $\chi 2$ (1) = 67.17, p <.001, and pointing, $\chi 2$ (2) = 9.14, p=.010, but the main effect of eye gaze was not significant, $\chi 2$ (2) = 1.33, p = .515. There was one significant interaction, between pointing and object location, $\chi 2$ (2) = 50.36, p<.001 (Figure 2.4). None of the interactions involving gaze were significant; pointing and gaze, $\chi 2$ (4) = 3.70, p=.448, verbal cue and gaze, $\chi 2$ (2) = 4.50, p=.106 and the three-way interaction between all factors, $\chi 2$ (4) = 6.94, p =.139.

As expected, participants were overall more likely to pick the word *this* (90% of the time) when the object was near the speaker compared to when the object was far from the speaker (*OR*=133, *SE*=79.6, *z*=8.20, *p*<.001). Similarly, participants were overall more likely to pick the word *this* when the agent was pointing at the near object as opposed to when they were pointing at the far object, *OR*=3.21, *SE*= 1.18, *z*=3.02, *p* =.008, although the contrast of trials with and without pointing did not yield any significant result (seeing proximal or distal pointing did not lead to any in demonstrative choice compared to its absence, *OR*=2.25, *SE*= .892, *z*=2.05, *p* =.121 and *OR*=.72, *SE*= .214, *z*=-1.10, *p* =.814 respectively.

The significant interaction between pointing and object location suggests two main findings (Figure 2.4). Firstly, combining pointing gesture with noun referring to one of the objects (congruently) did not confer an advantage over seeing the noun alone, either when both cues were targeting the proximal location, *OR*=3.19, *SE*= 1.44, *z*=2.57, *p*=.15 or the distal location, *OR*=1.75, *SE*= 0.64, *z*=1.53, *p*=1.00. In other words, the absence of a (congruent) point with a demonstrative did not lead to increased use of the default demonstrative form for the proximal location. Secondly, when the noun representing the proximal object was paired with distal (incongruent) pointing, participants were less likely to pick *this* compared to trials in which two cues were (congruently) targeting the proximal item, *OR*=0.09, *SE*= 0.04, *z*=-5.55, *p*<.001, or trials when there was no pointing, *OR*=0.297, *SE*= 0.10, *z*=-3.77, *p*=.003. Therefore, distal pointing affected demonstrative choice when the object location was proximal. However, there was no effect of pointing when the noun was representing the distal object; distal object location with (incongruent) proximal pointing did not lead to a significant increase in the likelihood of picking *this*, compared to trials when the two cues were congruently distal *OR*=0.91, *SE*= 0.38, *z*=-.23, *p*=1.00 or when there was no pointing, *OR*=1.59, *SE*= 0.70, *z*=1.05, *p*=1.00. This suggests that an inconsistent point leads to the use of the default term, *that*, but the effect of incongruence only occurs for the proximal location.

Figure 2.4: Interaction of pointing and object location. The error bars represent standard error.



2.1.4 Discussion

Demonstratives usually cooccur with gaze and gesture in naturalistic interactions (Todisco et al., 2020), especially when the object referent is proximal to the speaker (Bangeter, 2004; Cooperrider, 2016). In this study, we asked two fundamental questions regarding the interplay between these cues. First, we asked if synchronous gaze and gesture are *necessary* for perceived potential to use the proximal demonstrative term, consistent with the view that demonstratives and gestures are fully integrated into a single lexical representation (Terenghi, 2022). Accordingly, we tested if the neutral demonstrative (*that*) is chosen more in the absence of other congruent deictic cues compared to the presence of congruent cues. Second, we tested if a mismatch between cues affects the demonstrative choice by pitting object location, gesture and eye gaze cues against each other. We consider each of these goals in turn.

Pointing and object location were significant factors in choosing spatial demonstratives (dichotomously). Overall participants were more likely to choose the word *this* when either pointing or noun were targeting the proximal item and word *that* when they were targeting the distal item. However, with regard to our first goal, the results showed clearly that the presence of a congruent pointing gesture does not lead to increased decisiveness in demonstrative choice. This suggests that pointing is not mandatorily integrated into the representations for demonstratives, as has been suggested (Terenghi, 2022). Rather, the co-occurrence of demonstratives with pointing may afford other advantages rather than simply replicating deictic pointing across two modalities. While pointing directs the attention of an addressee to an object, demonstratives have a much wider function than merely directing attention to an object, as has been recognised in the recent literature on demonstratives (Diessel & Coventry, 2020)

We also wanted to see whether demonstrative choices are affected by mismatches between deictic cues, for example choosing one of the demonstratives more often than the other when pointing and/or eye gaze were incongruent. When the verbal cue was referring to the proximal item and pointing was distal, participants were more likely to pick the word *that* even though the noun in the two-word sentence referred to the proximal item. In contrast, no effect of mismatch was found when the noun denoted the distal object and pointing was proximal. We take this pattern of results to indicate that ambiguity due to a mismatch between two cues led participants to choose the word *that* with its possible definitive function when the object was located proximally. When the object was located distally, the use of the word *that* is already appropriate to refer to the distally located object, and the same term has a definite (non-spatial) function, hence the effect of a mismatch only when the object was located proximally.

Surprisingly, eye gaze did not have any significant effect in deciding a spatial demonstrative, although previously it is argued to be an important factor in modulating joint attention (Holler et al., 2012, 2014; Stukenbrock, 2020). In this study, the strength of pointing and object location as factors in demonstrative choices might have made the eye gaze redundant. In other words, in relation to the other two cues, eye gaze might have not had a noticeable effect on the demonstrative when the choice is dichotomous. Moreover, the role of eye gaze may be more relevant when a speaker and addressee are interacting, where it has been argued that monitoring the presence or absence of joint attention was important for demonstrative use (Caruana et al., 2021; Diessel, 2006; Hanna, Brennan & Savietta, 2020; Stukenbrock, 2020).

Overall, the results reveal two new findings. First, the presence of a gesture with a proximal target is not necessary in English for the use of the proximal term. While in some languages it has been argued that pointing is obligatory with demonstratives (Bohnemeyer, 2018; Hermann, 2018; Meira, 2018), this is not the case in English. Second, the results highlight the importance of consideration of mismatches between modalities and the conditions under which default or neutral demonstrative forms are used (Abashidze & Knoeferle, 2021; Ahn, 2017; Goldin-Meadow, 2014; Lepore & Ludwig, 2000). In Turkish it has been reported that there is a specific demonstrative form, *şu*, that is used when the addressee is not attending to the interaction (i.e. when the usual condition of joint attention for demonstrative use is flouted; Küntay & Özyürek, 2006). In languages that do not have such specific demonstratives to do this, neutral or default forms may be used under a broader range of conditions not limited to breaches of joint attention. Future work would do well to investigate how demonstratives are used in more naturalistic interactions under conditions of altered attention and incongruence of modalities.

2.2 Experiments 2, 3 & 4

2.2.1 Introduction

Spatial deixis is paradigmatic of communication and serves to manipulate an interlocutor's focus of attention towards a referent (Diessel, 2006; Diessel & Coventry, 2020; Kita, 2003; Levinson, 1983; Stukenbrock, 2015) and can be conveyed through the modalities of language use (e.g., the spatial demonstratives *this/that/here/there*), eye gaze and pointing gestures (Iverson et al., 1994; Liszkowski et al., 2012). Just as pointing projects from a speaker through an outstretched arm to the location of the (intended) referent, spatial demonstratives and eye gaze can be understood via the same underlying vectorial representation (Fricke, 2014; Iverson et al., 2000; Kita, 2003; Todisco, Guijarro-Fuentes, Collier & Coventry, 2020).

In practice, spatial demonstratives, eye gaze and gesture often work in tandem during deictic communication (Carpenter et al., 1998; Talmy, 2018). For example, Todisco et al. (2020) analysed the deictic communications of caregivers and infants during naturalistic storybook interactions, finding that infants use demonstratives, pointing and eye gaze synchronously in communication. As mentioned earlier, in some languages, it is in fact obligatory to point or gaze when using spatial demonstratives (e.g. Goemai, Hellwig, 2003; Kilivila, Senft, 2004; Yucatec, Bohnemeyer, 2018; Warao, Herrmann, 2018; Tiriyó, Meira, 2018). This begs the question as to why such deictic cues usually co-occur. Using multiple cues is more effortful than employing single cues, and therefore one can question if multiple cues confer greater decisiveness with which hearers are able to identify the speaker's intended referent than single cues alone. Spatial demonstratives, eye gaze and gesture are all assumed to be powerful deictic tools in communication, but their relative strength in deictic communication has yet to be understood.

Moreover, while one of the key functions of demonstratives is to direct the attention of a hearer to an object in space, just how effective demonstratives are as deictic cues has not been established.

The main objectives of the current work are to establish the relative importance of language, gesture and eye gaze as deictic cues guiding attention to a potential referent. In doing so we compare congruent and incongruent modalities in spatial reference choice, while also probing dynamic versus static signalling for reference. The end goal is to understand just how effective individual and multiple cues are in deictic reference, and in doing so to illuminate why deictic communication is invariably multimodal.

Prior to presenting a series of experiments manipulating deictic cues across modalities, I briefly recap the evidence for the importance of individual deictic cues before considering their relative importance.

2.2.1.1 Single and multiple (modal) deictic cues

Demonstratives

The function of spatial demonstratives – words including *this, that, here,* and *there* in English – has been much debated in recent years (see Diessel & Coventry, 2020 for a recent review). It is generally accepted that demonstratives serve two closely related functions, indicating spatial location, and coordinating speaker-listener (joint) attention (Diessel, 2006). With regards to the indication of spatial location, many have argued that demonstratives contrast distance, with a proximal term used for an object (referent) near a speaker (or in peripersonal space) and a distal term for an object far from a speaker (Anderson & Keenan, 1985; Bühler, 1934; Diessel, 2014).

As mentioned earlier, in English, it has been shown that *this* and *that* map onto a more precise peripersonal and extrapersonal space distinction (Caldano & Coventry, 2019; Coventry et al. 2008, 2014; see also Kemmerer, 1999). Coventry and colleagues (2008, 2014) systematically manipulated the distance an object was positioned in the sagittal plane on a table in front of participants and found that *this* was used significantly more when the object was reachable by the speaker while *that* was used more when the object was out of reach. Moreover, extending reach through the use of a tool led to a corresponding extension of the use of *this* to describe object location beyond the end of the hand to the end of the tool (Coventry et al., 2008).
So, *this* can be regarded as a cue for an addressee to look for an object within the peripersonal space of the speaker while *that* should draw attention to a referent distal from the speaker. Indeed, such mapping has been confirmed via event-related potentials when a heard demonstrative is incongruent with the expected proximal-distal mapping (Stevens & Zhang, 2013). While distance/reachability is important for demonstrative choice in English, it is also the case that demonstrative use across languages is affected by a wider range of parameters (Coventry, Griffiths & Hamilton, 2014; Diessel & Coventry, 2020; Levinson et al., 2018; Peeters, Krahmer & Maes, 2021).

Gestures and eye gaze as unimodal cues

Gestures can be defined as "visible action when it is used as an utterance or as part of an utterance" (Kendon, 2004, 7). Included in such gestures are pointing with the arm, or other body parts, but also shrugs and nods. Pointing is a ubiquitous form of indicating (Clark, 2003; Kendon, 2004; Cooperrider et al., 2021; Talmy, 2018). It appears to be universal across cultures (Eibl-Eisesfelt, 1989; Cooperrider et al. 2018) and (as briefly reviewed in Chapter 1) developmentally is among the earliest forms of communication with an evolutionary value in language (Bates, 1979; Capirci & Volterra, 2008; Liszkowski, 2012). Infants as young as 4-6 months are capable of covertly shifting their attention to the intended reference of a pointing hand (Bertenthal, Boyer & Harding, 2014). Moreover, pointing gestures are extremely frequent during spatial communication (Kita, 2003). Cooney, Brady and McKinney (2018) have shown that people are extremely accurate at judging the intended location when a conspecific is pointing with arm outstretched and index finger extended.

There is also much evidence for the power of eye gaze in directing attention. Sensitivity to eye gaze appears early in development, with evidence that infants at 12 months are able to accurately discriminate the object that someone is looking at (e.g., Phillips, Wellman & Spelke, 2002). Furthermore, multiple studies with adults show that people automatically shift attention as a function of eye gaze; when people observe someone looking in a particular direction, their attention is shifted to the same location in space (e.g., Driver et al., 1999; Edwards, Stephenson, Dalmaso & Bayliss, 2016; Friesen & Kingstone, 1998; Hietanen, 1999; Langton & Bruce, 1999). Such automaticity is reflected in studies showing that a gaze incongruent with a cued target location results in a cost responding to that location, even when participants are told with 100% certainty where the target will appear (Galfano, Dalmaso, Marzoli, Pavan, Coricelli & Casteli, 2012). In deictic communication, Hanna and Brennan (2007) have shown that participants, in a spontaneous dialogue task, use the gaze of their interlocutors to identify targets in a matching task before hearing an object name that disambiguated reference.

Cues in Combination

While language, gesture and eye gaze are powerful (individual) cues to intended reference, as debated in Chapter 1, regarding their relative contribution in deictic reference is less clear.

Several studies have broadly examined the relationship between language and gesture (e.g., Kelly, Healy, Özyürek & Holler, 2014; Kelly & Lee, 2012; Kelly, Özyürek & Maris, 2010), and the relationship between language, gesture and gaze (Holler, Schubotz, Kelly, Hagoort, Schuetze & Özyürek, 2014). For example, Holler et al. (2014) presented videos of speech only or speech and gesture object-related messages (nouns), with the speaker either gazing at the participants or sideways to another (out of view) person. They found that participants addressed directly in the speech-only condition were quicker to select the correct object than those in the unaddressed speech-only condition. However, participants in the speech plus (iconic) gesture condition when unaddressed performed as well as those in addressed conditions. Holler et al. argues that speech processing suffers when speaker's gaze is averted but gesture processing does not, therefore facilitating overall comprehension (the *cross-modal priming hypothesis*). While these results suggest the interplay between language, gaze, and gesture in object-noun processed subsequent recall, only a small number have examined the relationship between deictic pointing and linguistic spatial deixis on the one hand and gaze and linguistic spatial deixis on the other.

A few studies have examined the interaction between demonstratives and pointing gestures during interactive spatial description tasks. As described in Chapter 1, Bangerter (2004) has shown that, when participants gesture to describe object location, they employ short deictic descriptions, but when they do not gesture, descriptions are longer. Additionally, Bangerter found that gesture is more frequent when participants describe proximal locations rather than distal locations (see also Cooperrider, 2016), and Piwek, Beun and Cremers (2008) reported that distals were preferred when not pointing over proximals. So, while it has been often assumed that demonstratives and pointing are two sides of the same coin, patterns of cooccurrence are likely to be more nuanced.

To our knowledge, only four studies have directly manipulated demonstrative use and gesture or gaze. The only study to directly compare the effects of gesture plus deictic description to the deictic description or pointing alone was conducted by Bangerter and Lowerse (2005). Participants viewed video clips of a person describing and/or pointing to an array of objects on a computer monitor. Accuracy rates for correct identification of (location) targets by participants were similarly high when the person in the video gave a spatial description alone or pointed alone, but there was further improvement in target identification in the pointing plus spatial description condition. Using a rather different method, Chieffi, Secchi and Gentilucci (2009) asked participants to read words aloud, with words on cards placed either close or far from them, while simultaneously pointing at their own body when the card was near and pointing towards a far location then the card was placed far. They found that the content of the word affected both voice spectra and pointing kinematics, suggesting a bidirectional interaction between speech and gesture production systems.

Our goal in the experiments that follow was to examine the extent to which language, gesture and gaze guide attention to a referent in deictic communication for the first time. In a series of online experiments, participants saw images of a person (agent) referring to one of two objects placed sagittally. The task for participants was to identify the object they thought the agent was referring to (Figure 2.7). Both experiments manipulated whether, in addition to language, the agent only pointed, only gazed, or pointed and gazed at one of the two objects placed in front of them (Figure 2.2 and 2.3), while manipulating the congruence of verbal (demonstratives) and non-verbal cues (pointing and gaze). We had two primary goals. First, we wanted to test the relative strengths of individual cues versus multiple congruent cues on reference choice. Given that deictic communication usually involves multiple cues, we predicted that multiple cues (e.g., Figure 2.2a) might be more effective cues to a reference than single cues (e.g., Figures 2.2b, 2c). This would be consistent with the findings of Bangerter and Lowerse (2005) with respect to pointing and spatial description. Second, to investigate the relative strengths of cues we wanted to pit cues against each other by manipulating their congruency (e.g., Figure 2.2e). Following evidence that iconic gestures overcome gaze aversion (Holler et al., 2014), we predicted that pointing would dominate reference choice when the gaze is on a different referent. In addition to the primary goals, a secondary goal was also to compare cues and combinations of cues in cases where the non-verbal cues are static (shown in photographs) versus dynamic (shown in videos).

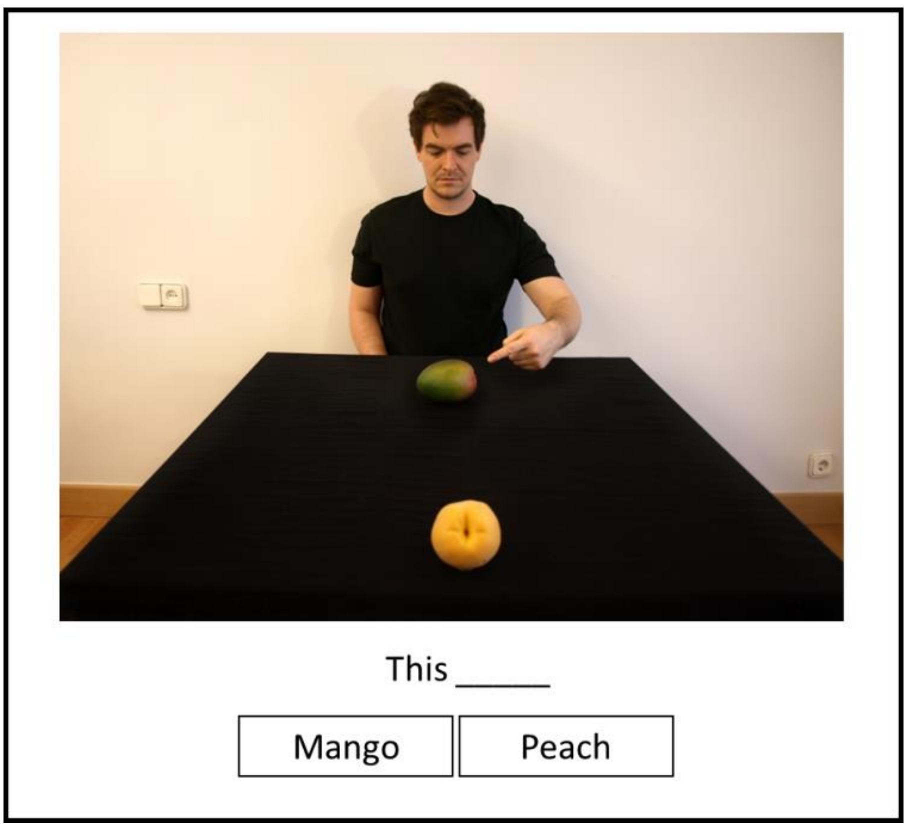
2.2.2 Experiment 2 (Pilot Study)

2.2.2.1 Methodology

Participants: Participants were University of East Anglia students participating in the study for course credit. Power analysis carried prior to recruitment using G*Power version (3.1) (Faul, Erdfelder, Buchner & Lang, 2009) to detect a medium effect (f = 0.25) at a significance criterion of α=.05 indicated a sample size of 36 participants for this experiment. Participants who did not speak English as their first language were excluded from analyses, were a total of 34 participants after exclusion (Female: 30, M_{age}=20.7, SD_{age}=2.95).

Stimuli, Design and Procedure: The study was conducted online through Gorilla (Gorilla.sc) experimental platform. Participants saw images of a person (agent) coupled with a spatial demonstrative. The agent (male) was pointing (left arm) and gazing towards one of the two potential referents, one positioned proximally (near) one position distally (far) relative to the agent (Figure 2.2). The experiment employed a (3 x 3 x 2) full factorial repeated measures design with three levels of pointing, eye gaze and spatial demonstratives. The pointing arm varied in three levels, the agent would either point at the proximal item, distal item or he would rest his arm on this leg (no pointing). Eye gaze also varied in three levels, targeting either proximally or distally, and agent also had his eyes closed (no gaze). The images were coupled with a spatial demonstrative, either *this* or *that*, for every level of pointing and eye gaze. The task was to decide which one of the referents the agent was targeting to complete a two-word sentence (see Figure 2.3 for the design). The options were presented under the two-word sentence with a missing word and the lateral position in which these two options were presented was randomised. The response options and the stimuli were presented simultaneously (Figure 2.5). Participants would proceed to next trial once they chose a noun.

Figure 2.5: Example of the experimental stimuli.



Same as Experiment 1, there were 18 possible combinations of the levels of pointing (proximal item, distal item, and arm not visible) eye gaze (proximal item, distal item and eyes closed) and spatial demonstratives (*this* or *that*). Trials consisted of a combination of the three variables. There was a total of 72 trials per participant, divided into four blocks. Therefore, there were four data points per level of the design. The trials within a block and order of blocks presented to each participant were randomised. At the end of four blocks, there was a fifth block of gaze detection task with six trials to see how accurate participants were in detecting gaze from the images.

Data preparation: For each time a participant picked the proximal item, they were given one (1) point and zero (0) for the picking of the distal item. The data were analysed in terms of participants' picking of the proximal item. There was a total of 72 data points per participant.

Analysis model: Statistical analysis was performed in Jamovi (version 2.2; The jamovi project, 2021) and analysed using Generalized Mixed Models (Gallucci, 2019). Outcome was computed as the likelihood of participants choosing the proximal object. Due to the binomial nature of the outcome (1 and 0), the model had a binominal distribution and logit link function. We included random coefficients with participants as intercepts, predictors and predictor interactions as random slopes, as long as the model converged, and model fit was not singular. The random effects structure was formed considering the convergence and theoretical relevance (Barr, Levy, Scheepers & Tily, 2013). The final random coefficient structure included demonstrative and pointing. Fixed effect factors included pointing, eye gaze, verbal cue, and all possible interactions. Models were estimated using log likelihood (odds ratio, OR). The main effect and interaction of the predictors reported in text are from fixed effects omnibus tests. The fixed effect coefficients were assessed using Wald's *Z* statistics. Fixed effects parameter estimates reported were coded with Helmert contrast (Table 7.2). For interpretability, inverse odds are reported when the OR is less than 1 (1/OR). Unless stated otherwise, the graphs and percentages reported are based on estimated marginal means. Significant interactions and contrasts within a factor were then further investigated with posthoc analyses with Bonferroni correction.

2.2.2.2 Results

We first examined the gaze detection task to ensure participants were able to identify where the person in the image was looking. The mean gaze detection task score was 5.68 (*SD*=.59) out of a possible perfect score of 6.

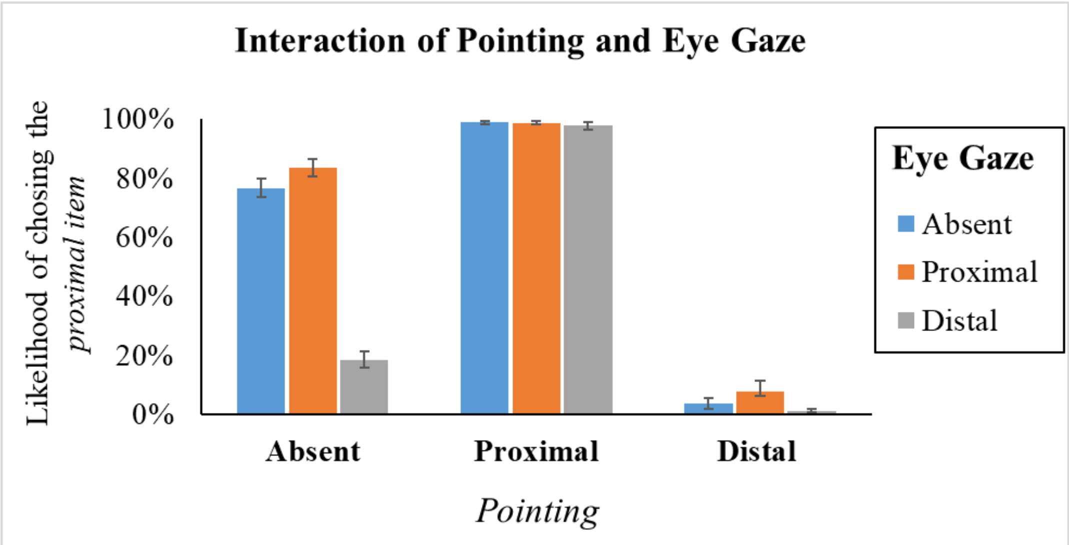
The model testing the main effect and interactions (Conditional $R^2 = 0.82$) of pointing, gaze and seeing a spatial demonstrative indicated a significant main effect of pointing, $\chi 2$ (2) = 78.17, $p<.001$ and eye gaze $\chi 2$ (2) = 94.37, $p<.001$ while spatial demonstratives did not have a significant main effect, $\chi 2$ (1) = 3.10, p=.08. Overall, participants picked the proximal item 54.2% of the time.

As expected, participants were more likely to pick the proximal item when the agent was pointing towards it compared to the trials in which agent was pointing towards the farther referent, *OR*=2018.10 *SE*=1739.30, *z*=8.83, $p<.001$ or when there was no pointing (*M*=.61, *SE*=.03), *OR*=42.91, *SE*= 21.17, *z*=7.62, $p<.001$. Similarly, distal pointing led to a decrease in the likelihood of choosing the proximal item compared to trials without pointing, *OR*=50, *SE*= 0.01, *z*=-7.32, $p<.001$. Overall, when the agent was pointing towards the proximal item, participants picked the proximal object 98% (*SE*=.008) of the time and when pointing was distal, only 3% (*SE*=.01).

Eye gaze also was a significant factor in choosing a referent. In a similar trend to the effect of pointing, participants were more likely to pick the proximal referent (*M*=.77,*SE*=.04) in trials which eye gaze was proximal compared to trials with distal eye gaze (*M*=.33, *SE*=.05), *OR*=6.78, *SE*=1.39, *z*=9.34, *p*<.001. Furthermore, participants were less likely to pick the proximal item when eye gaze was targeting the farther object, *OR*=4.55, *SE*= 0.04, *z*=-7.43, *p*<.001 compared to trials without any eye gaze.

Pointing and eye gaze also had a significant interaction $\chi2$ (4) = 0.55, p<.001 (Figure 2.6). Overall, pointing was the most dominant cue. When pointing was proximal there was no effect of gaze, in all contrasts *p*>.005. However, when pointing was distal there was an effect of proximal gaze, *OR*=7.39 *SE*= 2.79, *z*=5.31, *p*<.001 with 6 % increase in likelihood of choosing the proximal item compared to trials in which both cues are targeting the farther item. In the absence of pointing, proximal gaze led to a higher chance of picking the proximal item compared to distal gaze, *OR*=22.74, *SE*= 6.05, *z*=11.74, *p*<.001.

Figure 2.6: Interaction of pointing and eye gaze. Error bars represent standard error.



However, surprisingly, the absence of both cues still led %77 (*SE*= .03) likelihood to choose the proximal item, but this effect was not as strong as seeing two cues congruently (at 99%), *OR*=24.77, *SE*= 14.09, *z*=5.64, *p*<.001.

2.2.2.3 Summary

In this experiment, we manipulated the demonstrative seen, pointing and eye gaze information cued by the agent. The results suggested a significant role of pointing and eye gaze but surprisingly not of spatial demonstrative seen. Pointing and eye gaze also had a significant interaction. The interaction indicated firstly that seeing proximal pointing alone in absence of gaze was a sufficient cue for reference. However, when pointing is distal, there is an effect of proximal eye gaze, therefore when incongruent to pointing, gaze would lead to a slightly increased chance of picking the proximal item. Interestingly, the participants were more inclined to pick the proximal item overall, even in the absence of the two cues (although not as highly likely as when pointing was proximal), with or without (congruent or incongruent) eye gaze. This could be due to one or both these two factors. Firstly, previous research indicated a predisposition towards proximal referent in ambiguity, where in absence of both cues, participants resorted in picking the object closer to the agent, since it is more accessible/desirable to the agent. Secondly, although demonstratives did not show a statistically significant effect, they were still present. Seeing that participants were still highly to picking the proximal item when pointing and gaze were absent could implying a potential effect of demonstrative which simply did not reach to significance in this experiment due to, for example, methodological reasons. Potentially, demonstratives did not capture as much attention as pointing or eye gaze in the computer screen.

2.2.3 Experiment 3

The goal of Experiment 3 was to examine the impact of single versus multiple deictic cues and the congruence of cues on referent choice (Figure 2.7). Participants saw images of two objects on a table in front of the agent using language (*this, that*) with or without (congruent or incongruent) pointing and eye gaze to refer to one of the objects. Participants had to decide which of the two objects (a near object and a far object) they thought the agent was referring to.

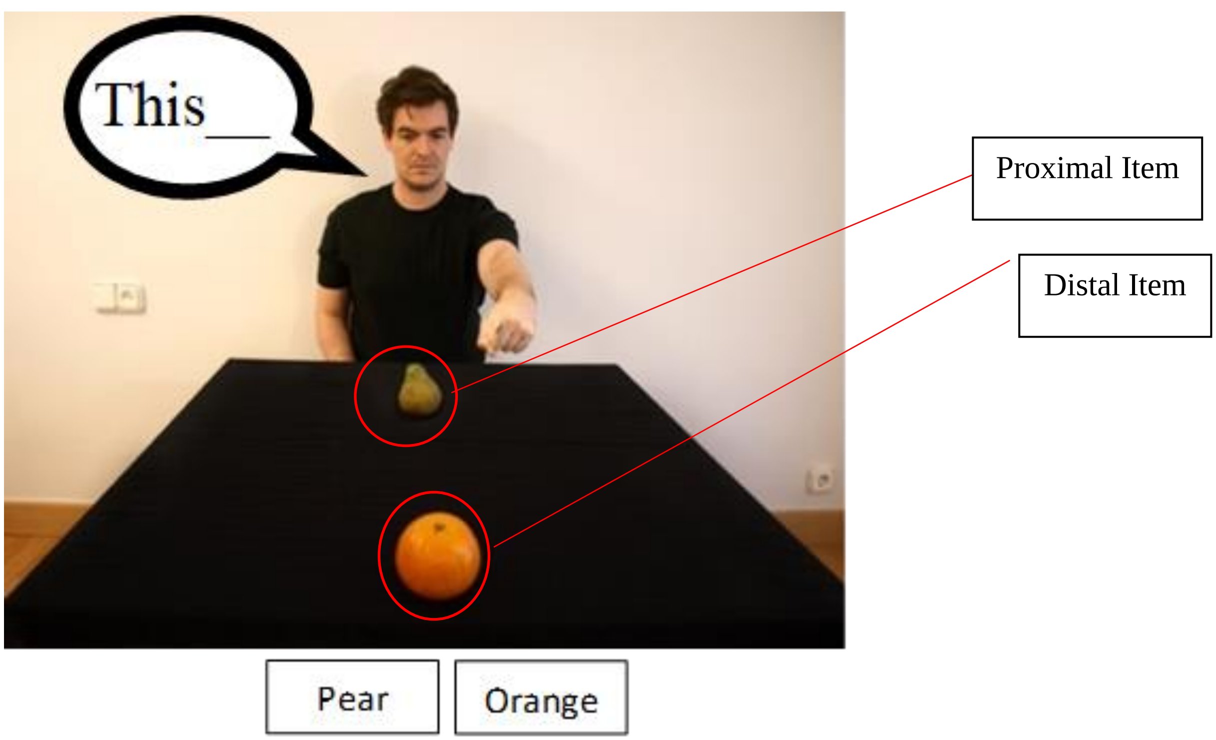
2.2.3.1 Methodology

Participants: Participants were University of East Anglia students participating in the study for course credit. Power analysis carried out prior to recruitment using G*Power (version 3.1) (Faul, Erdfelder, Buchner & Lang, 2009) to detect a medium effect (f = 0.25) at a significance criterion of $\alpha = .05$ indicated a sample size of 36 participants for this experiment. Participants who did not speak English as their first language were excluded from analyses, leaving a total number of 33 participants (M_{age} = 21.9, SD_{age}=8.18; 29 females) after exclusion.

Stimuli and Design: The experiment was presented on the online experimentation platform Gorilla (Gorilla.sc). Stimuli on each trial consisted of images of the agent (male) sitting behind a table, pointing, and gazing at one of two items placed on a sagittal plane; one positioned proximally (close) and one distally (far) relative to the agent. The agent always pointed with his left arm. The task was to complete a two-word sentence with a missing word, placed in a speech bubble (Figure 2.7). The position of the speech bubble was counterbalanced where half of the participants saw the speech bubble to the left of the agent, and half on the right. Response options were positioned under the image and participants were asked to click and select from one of the two options (proximal or distal item). The presentation order of the object choices was counterbalanced (with the proximal item appearing on the left 50% of the time).

There were three main manipulations in the experiment: the spatial demonstrative (verbal cue) presented in the two-word sentence (either *this*, targeting proximal item or *that*, targeting the distal item), pointing arm, and eye gaze, in a 2 x 3 x 3 repeated measures design. Similar to previous experiments reported, the pointing arm varied across three levels; no pointing (the arms of the agent resting on his leg), pointing to the item closer or farther relative to the agent. The eye gaze of the agent also varied across three levels; no gaze (eyes closed), gaze targeting the proximal or distal object (Figure 2.3).

Figure 2.7: Example of the task.



At the beginning of the experiment, participants were introduced to the task. Once informed consent was given, participants completed an example trial, making it clear that their task was to select the missing word the person in the picture would utter according to what object they thought the person was referring to, via a button option beneath the image (Figure 2.7). Before the example trial, there was a familiarisation stage during which participants saw the items they would see through the trials and their associated labels (all fruits with similar rounded shapes). The position of the items on the table was counterbalanced. Each item appeared in each position (near and far) and combinations were unique for each trial.

Each participant completed four blocks with 18 trials in each block, totalling 72 experimental trials, with four data points for each level of the design per participant. The order of blocks and trials within each block were randomised between participants. At the end of the task, participants completed a gaze detection task (with six trials) to check that they were accurate in detecting the gaze of the agent.

Data preparation: Every time a participant picked the proximal item, they were given a score of one (1) and zero (0) for the distal item. The data were then analysed in terms of the likelihood of picking the proximal item.

Analysis Model: Statistical analyses were performed in Jamovi (version 2.2; The jamovi project, 2021) and analysed using Generalized Mixed Models (Gallucci, 2019). Due to the binomial nature of the outcome (1 and 0), models had a binominal distribution and logit link function. The outcome was computed as the likelihood of participants choosing the proximal object. We included random coefficients with participants as intercepts, predictors and predictor interactions as random slopes, as long as the model converged, and the model fit was not singular. The random effects structure was formed considering the convergence and theoretical relevance (Barr, Levy, Scheepers & Tily, 2013). The final random coefficient structure included participants and demonstrative seen. Fixed effect factors included pointing, eye gaze, verbal cue, and all possible interactions. Models were estimated using log likelihood (odds ratio, OR). The main effect and interaction of the predictors reported in the text are from fixed effects omnibus tests. The fixed effect coefficients were assessed using Wald's Z statistics (Table 7.3). Fixed effects parameter estimates are reported in tables and all factors were coded with Helmert contrast (these are not reported in the text but in tables in the appendix)). Again for interpretability, inverse odds are reported when the OR is less than 1 (1/OR). Unless stated otherwise, the graphs and percentages reported are based on estimated marginal means. Significant interactions and contrasts within a factor were then further investigated with post hoc analyses with Bonferroni correction.

2.2.3.2 Results

We first examined the gaze detection task to ensure participants were able to identify where the person in the image was looking. The mean gaze detection task score was 5.18 (*SD*=1.16) out of a possible perfect score of 6.

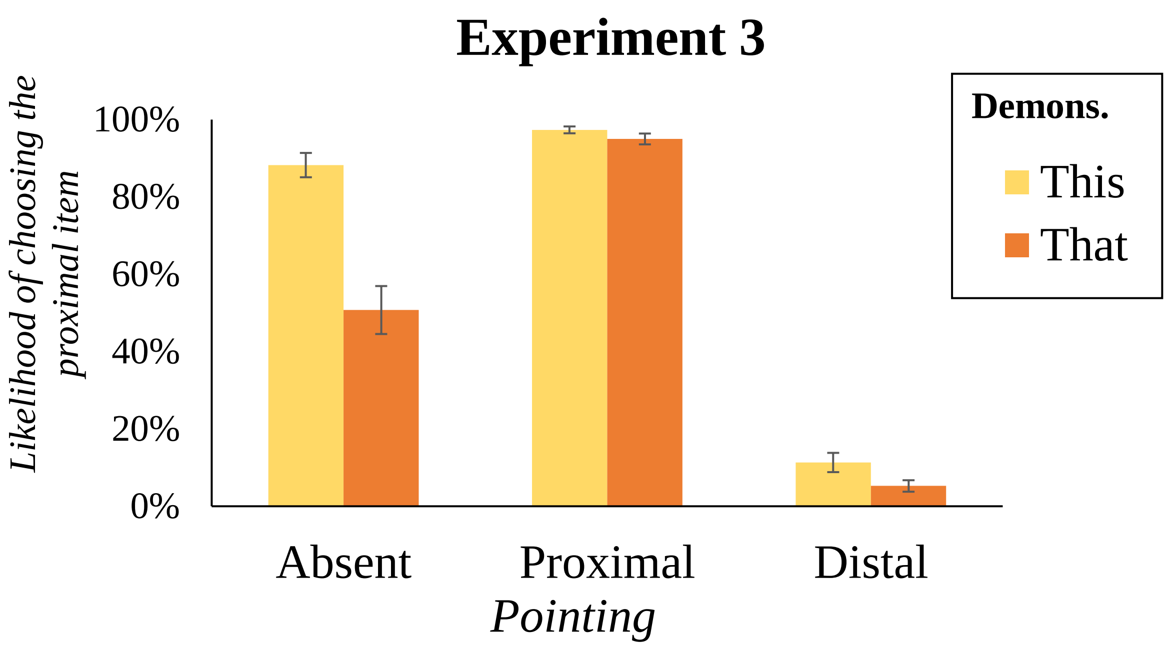
The model indicated significant effects of all three factors (Conditional R^2= .735), pointing, $\chi2$ (2) = 528.98, p <.001, eye gaze, $\chi2$ (2) = 146.19, p <.001 and spatial demonstratives, $\chi2(1)$ =10.33, p = .001, and significant interactions between pointing and gaze, $\chi2$ (4) = 107.89, p <.001 and pointing and demonstrative seen, $\chi2$ (2) = 10.29, p =.006. There was no significant interaction between demonstratives and eye gaze, $\chi2$ (2) = .78, p =.67, and the three-way interaction was also not significant, $\chi2$ (4) = .88, p =.93.

As expected, the nearer object was selected as referent when *this* was presented more than when *that* was presented, (OR=3.19, SE=1.15 z=3.21, p <.001), with participants selecting the nearer object 77% of the time when they saw *this* and 51% of the time when they saw *that*. Again, as expected, participants were more likely to pick the nearer object when the agent was pointing at it as opposed to when they were pointing at the far object, OR=313.31, SE= 80.43, z=22.39, p <.001, with the selection of the nearer object when the agent was pointing at it 96% of the time, and only 8% of the time when they were pointing at the far object. Similarly, when the agent was pointing at the nearer object participants were around 10 times more likely to pick the proximal item than when there was no pointing OR=9.48, SE=2.29, z=9.33, p <.001. And again, as expected, the overall choice of referent was affected by the object the agent was looking at. When the agent was gazing at the proximal object, participants selected it 83% of the time. On the other hand, when he gazed at the distal object, participants selected the proximal object only 21% of the time. Furthermore, participants were 9 times less likely to select the proximal object when the gaze was on the distal item compared to conditions when there was no gaze, OR=9.09, SE=.02, z=-9.58, p <.001 and 12 times more likely to select the proximal object when the imaged person was gazing at the proximal object compared to when gazing at the distal object, OR=12.88, SE= 2.92, z=11.28, p <.001. The contrast of proximal gaze with its absence did not yield significance, OR = 1.36, z=1.39, p =.49.

As mentioned above, there were also two significant interactions: between pointing and spatial demonstrative seen, and between pointing and eye gaze. The interaction between demonstrative and pointing (see Figure 2.8) reveals two findings. First, the choice of referent was more decisive when demonstrative was combined with (congruent) pointing. When participants saw *this* in combination with proximal pointing, they picked the proximal item 97%

of the time. When they saw *that* with distal pointing, they picked the proximal item only 1% of the time. Compared to trials with no pointing, participants were more likely to choose the closer item when *this* was combined with proximal pointing, *OR*=4.89, SE= 1.83, *z*=4.25, *p* <.001, and in trials when distal pointing was combined with seeing the word *that*, participants were 20 times less likely to choose the proximal item compared to trials when participants only saw *that*, *OR*=20, *SE*= .02, *z*=-9.43, *p* <.001.

Figure 2.8: Interaction between spatial demonstratives and pointing in Experiments 3. Error bars indicate standard error. The values are estimated marginal means taken from the statistical model.

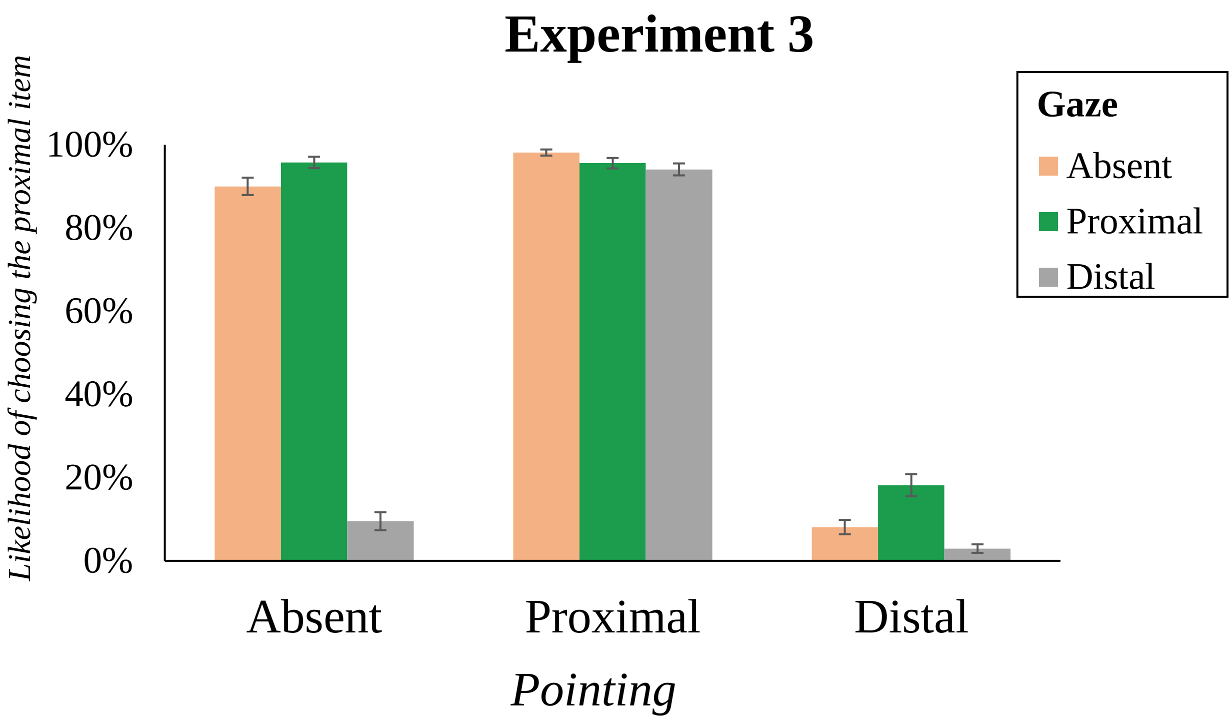


Second, when pointing and demonstrative are incongruent, the choice of the referent is determined by pointing and not by demonstrative. When a distal verbal cue (*that*) was paired with proximal pointing, participants were more likely to pick the proximal item (95% of the time) compared to trials in which both demonstratives and pointing referred congruently to the distal position, *OR*=342.07, SE= 124.27, *z*=16.06, *p* <.001, or trials when there was a distal demonstrative in the absence of pointing (51% of the time), *OR*=18.38, SE= 5.60, *z*=9.56 *p* <.001. When the proximal demonstrative, *this*, was paired with distal pointing, participants were less likely to pick the nearer object (11% of the time) compared to trials when both demonstratives and pointing target the near item congruently, *OR*=333, SE=001, *z*=-15.63, *p*<.001, or trials when there was a proximal demonstrative in the absence of pointing (88% of the time), *OR*=50, SE= 0.005, *z*=-12.86, *p* <.001.

There was also a significant interaction between pointing and eye gaze (Figure 2.9). In the absence of pointing, proximal gaze led to a 95% chance of picking the proximal referent compared to a 10% chance of picking the proximal referent when the gaze was distal, *OR*=215.03, SE= 89.10, *z*=12.96, *p* <.001. There was also an effect of gaze when pointing is distal, with more frequent choice of the far referent with (congruent) distal gaze (only 3% choice of the proximal object) compared to proximal gaze (at 18%), *OR*=7.32, SE= 2.80, *z*=5.21, *p* <.001. Critically, when there is a proximal point, there is no effect of gaze, *OR*=1.35, *SE*=0.27, *z*=-0.82, p=1.00. In other words, a proximal point appears to be a dominant cue to referent choice. Congruent proximal or distal pointing and gaze did not lead to a significant shift in choice of referent compared to pointing alone without gaze (*OR*=2.94, SE= 0.14, *z*=-2.63, *p*=.31 and *OR*=2.44, SE= 0.20, *z*=-1.87, *p* =1, respectively) or gaze alone without pointing (compared to trials with both pointing and gaze), *OR*=1.04, SE= 0.41, z=-0.11, p=1.00 and *OR*=3.45, SE= 0.12, z=-2.96, p=.11. Therefore, pointing is the dominant cue to referent choice.

Figure 2.9: Interaction of pointing and eye gaze in Experiments 3.

The values are estimated marginal means taken from the statistical model. Error bars represent standard error.



2.2.3.3 Discussion

Consistent with past studies, participants' choice of referent was influenced by all individual deictic cues, with the main effects of language, gesture and gaze. Results support previous literature showing that demonstratives are mapped onto perceptual space, as participants were more likely to pick the object within the speaker's reach when the demonstrative *this* was uttered by the speaker, and the farther referent when *that* was uttered (Anderson & Keenan, 1985;

Caldano & Coventry, 2019; Coventry et al. 2008, 2014). Regarding the role of gesture in deictic communication, pointing towards the nearer item led to an increased likelihood of picking the proximal item as referent, compared to trials without any pointing or with distal pointing. In relation to gaze as a cue, it was only effective when targeted towards the farther item, with participants more likely to pick the distal item compared to the no gaze condition (Proximal gaze did not affect referent choice compared to trials without gaze).

However, the pattern of interactions offers new insights into the interplay between deictic cues in referent selection. We asked if more deictic cues make it increasingly clear which object the speaker is referring to. Consistent with the results of Bangerter and Lowerse (2005), demonstrative plus (congruent) pointing led to more decisive referent choice by participants compared to language alone, both for proximal and distal deictic cues. However, the absence of an interaction between demonstrative and gaze suggests that congruent gaze does not increase the decisiveness of reference choice relative to language alone.

The pattern of interactions also shows the relative strengths of deictic cues when they are in conflict (incongruent cues). When language and pointing are incongruent, pointing dominates referent choice, overriding any influence of demonstrative. And when gaze and pointing conflict, gaze affects referent choice when pointing is distal but does not affect referent choice in the presence of a proximal point, partially consistent with the previous findings of Holler et al. (2014). Overall, while individual cues affect reference choice, pointing is the strongest predictor of reference in the data when pitted against demonstratives and eye gaze.

One might be tempted to conclude that pointing is the dominant cue, with the other non-verbal cue – gaze – in second place, with demonstratives the weakest cue to reference. However, demonstratives are usually heard/spoken in exophoric reference, and it is possible that participants in the first Experiment did not attend to the word in the speech bubble as much as they might do when they hear a demonstrative (relative to non-verbal cues). Moreover, using still images omits information about the dynamics of gaze and pointing that may influence action perception and henceforth behaviour (Haxby, Goddini & Nastase, 2020; Welke & Vessel, 2022). For these reasons, Experiment 4 attempted to replicate the findings of Experiment 1 using both dynamic and static images, with language presented auditorily.

2.2.4 Experiment 4

In Experiment 4, participants saw images (static trials) and 3-second videos (dynamic trials) of a person pointing, gazing, and verbally labelling ("this one" or "that one") one of the two potential referents. In contrast to Experiment 3, both images and videos were coupled with audio clips.

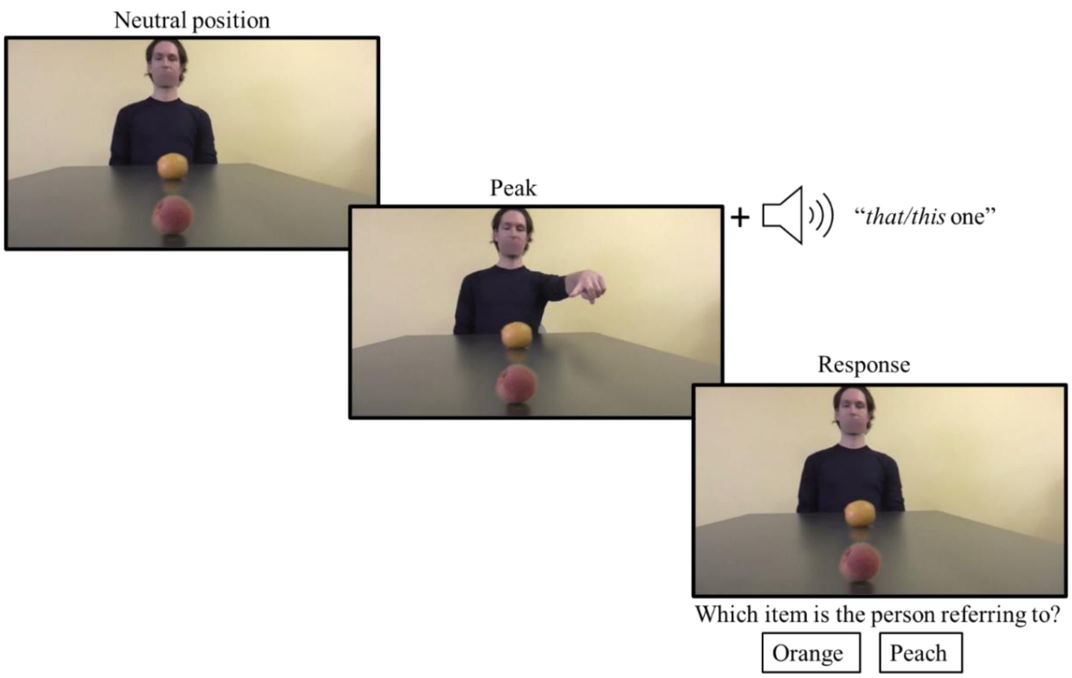
2.2.4.1 Methodology

The design for Experiment 4 comprised 36 levels; adding "dynamic" and "static" conditions doubled the design compared to Experiment 1 (the rest of the factors/levels were the same as in the previous experiment). Participants saw 18 videos and 18 stills per block, with a total of 36 trials in each block. As in the previous experiment, there were four blocks of experimental trials and a block of gaze detection trials. There was a total of 12 gaze detection trials: 6 with images and 6 with videos.

The videos in the dynamic condition started with the agent gazing straight ahead, arm resting on his leg. Then, he pointed and looked at one of the two potential referents, labelling the item with "this one" or "that one" (audio clips) and retracted his arm and diverted his eye gaze to the neutral position (Figure 2.10). The images consisted of stills from the videos at the peak of labelling where the recorded person was pointing, gazing, and verbally labelling one of the two possible referents. Participants were asked to answer, “Which item is the person referring to?” by picking the name of one of the two referents. The demonstratives were not written but only heard.

Figure 2.10: Example stimulus from Experiment 4, static condition.

The trials started with person gazing straight, followed by an image with the agent pointing and/or gazing coupled with audio, “*this*/*that* one”. The response option was presented in the neutral position. The mouth of imaged person was masked to avoid extra information being conveyed.



Participants: Participants were a combination of students at the University of East Anglia (*N*=6), completing the experiment for course credits, and participants from the Prolific recruitment platform (*N*=67), who were compensated for their time. There was a total of 73 participants (female=46; M_{age} = 22.6, SD_{age}=3.45), doubling the number of participants from the previous experiment as the design is now doubled with an added level of stimuli type. All participants were native speakers of English.

Data preparation: Every time a participant picked the proximal item, they were given a score of one (1) and a score of zero (0) if they selected the distal item. There was a total of 114 data points per participant.

Analysis Model: The same analysis model (generalized mixed model) as Experiment 3 was used but with the addition of stimuli type to the model. In the final model, random components included participants as intercept, demonstratives, pointing and gaze. Random components were kept to maximal as long as the model converged, and the model fit was not singular (Table 7.4 for full model,7.5 for static trials and 7.6 for dynamic trials).

2.2.4.2 Results

For the gaze detection trials, the highest possible score for the gaze detection task was 12, with a mean score of 9.86 overall (*SD*=1.98Participants were better at detecting gaze in dynamic trials (*M*=5.30, *SD*=.95) compared to static trials (*M*=4.56, *SD*=1.31), *t*(72)=5.47, *p*<.001.

The model testing the effects of stimuli type (videos, stills), pointing, eye gaze and spatial demonstrative (Conditional R^2= 0.84) produced significant main effects of spatial demonstrative, $\chi2$ (1) = 61.54, *p* <.001, pointing, $\chi2$ (2) = 543.42, *p* <.001 and eye gaze, $\chi2$ (2) = 43.26, *p* <.001. There were significant interactions between pointing and demonstrative, $\chi2$ (2) = 48.26, *p* <.001, and between pointing and gaze, $\chi2$ (4) = 162.82, *p* <.001. In the full model, there was also a significant interaction between eye gaze and demonstratives, $\chi2$ (2) = 2.43, *p* =.30. In addition, however, there were also a number of effects/interactions involving stimuli type (static versus dynamic); a main effect of stimuli type $\chi2$ (1) = 16.56, *p* <.001, and interactions between stimuli type and pointing $\chi2$ (2) = 10.04=5, *p* <.001, stimuli type and gaze $\chi2$ (2) = 22.48, *p* <.001, and significant three-way interactions between stimuli type, pointing and eye gaze, $\chi 2$ (4) = 12.15, *p* =.002, stimuli type, pointing and demonstratives $\chi 2$ (2) = 9.71, *p* =.008 and stimuli type, eye gaze and demonstratives, $\chi 2$ (2) = 7.96, *p* =.02.

The interaction of pointing and demonstrative replicated the results of Experiment 3 However, there was a slight difference in Experiment 3 regarding the interaction of pointing and eye gaze. In Experiment 4, congruent pointing and eye gaze targeting the distal position conferred an advantage over trials without pointing, *OR*=32.7, *SE*=.01, *z*=-10.70, *p*<.001.

Given the multiple interactions involving dynamic and static stimuli, to further understand how the nature of the stimuli affects referent choice we ran separate analyses for the two stimulus types (see appendix). The results for the static stimuli largely replicate the results from Experiment 3, with significant main effects as before of demonstrative, $\chi 2$ (1) = 54.54, *p* <.001, pointing $\chi 2$ (2) = 410.06, *p* <.001, and gaze, $\chi 2$ (2) = 32.57, *p* <.001, and interactions between demonstrative and pointing, $\chi 2$ (2) = 16.01, *p* <.001 and between pointing and gaze, $\chi 2$ (4) = 96.58, *p* <.001, all yielding the same patterns as Experiment 3 (Figures 2.11 and 2.12).

Figure 2.11: Interaction of pointing and demonstratives in Experiment 4 static image trials.

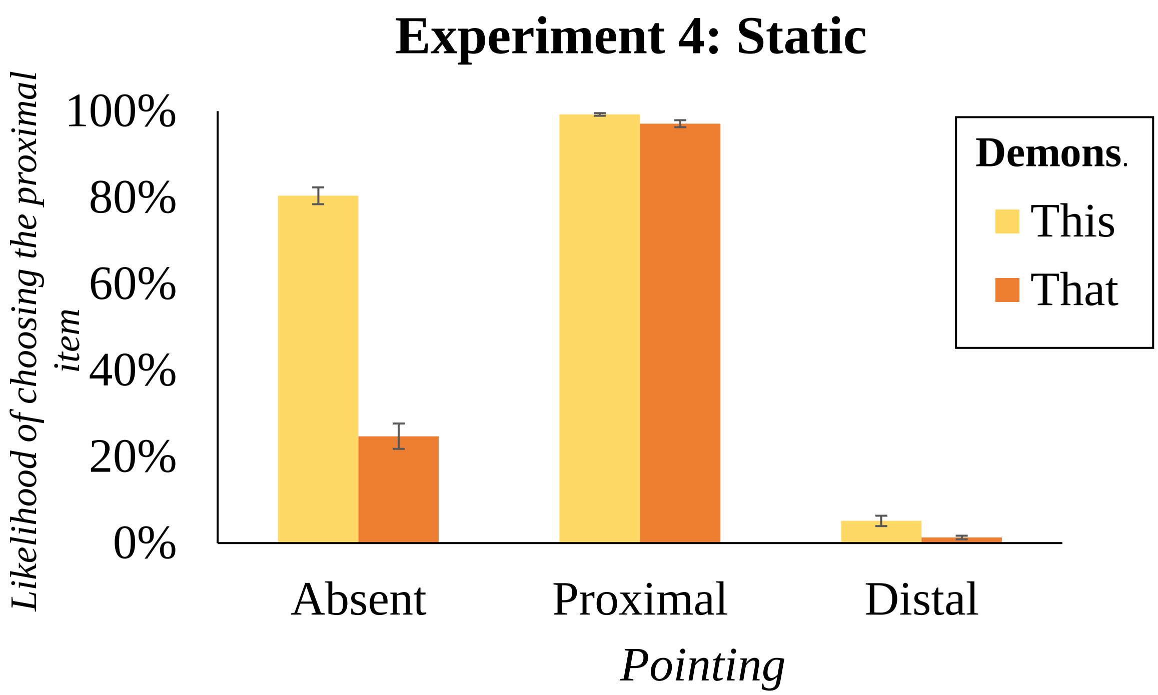
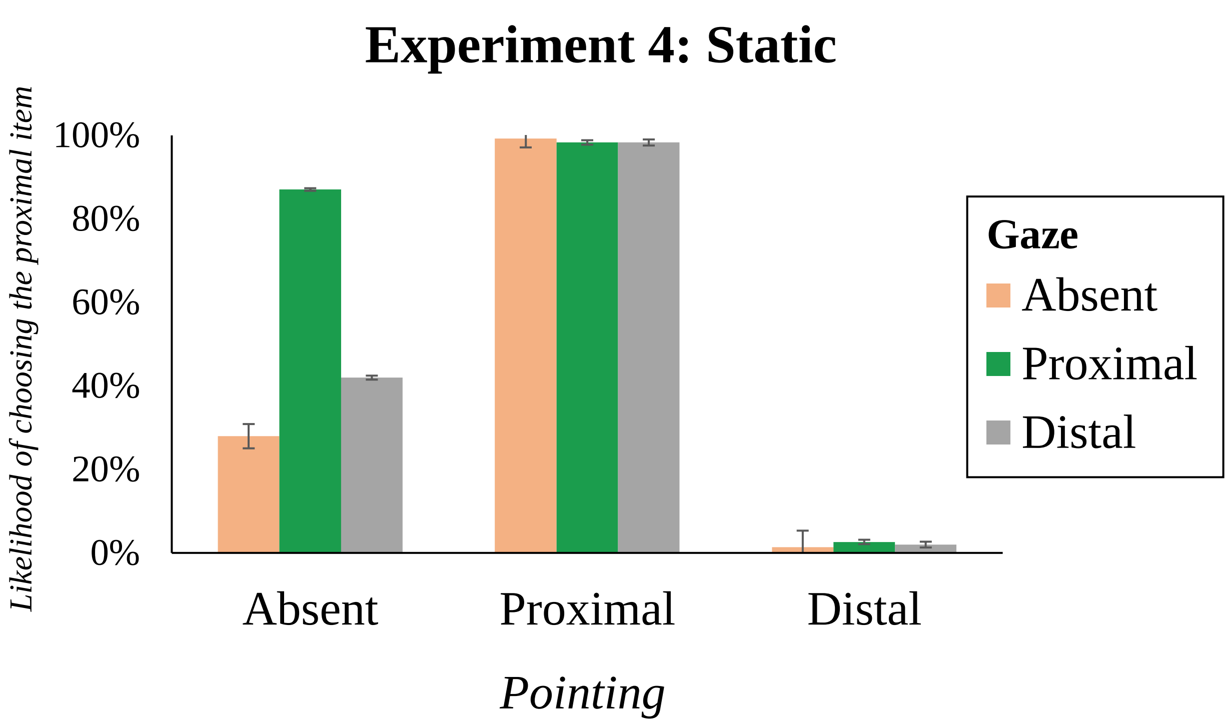
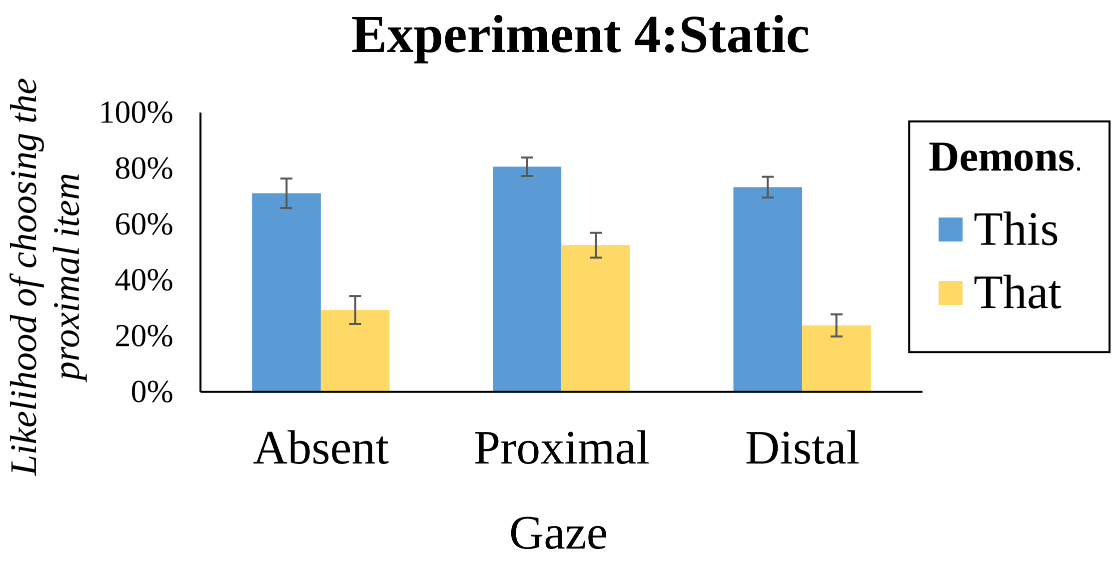


Figure 2.12: Interaction of pointing and gaze in Experiment 4 static image trials.



In addition, however, the static stimuli in this experiment also produced a significant interaction between demonstratives and eye gaze, $\chi2$ (2) = 7.22, p =.03 (Figure 2.13). Comparing the effect of demonstrative alone to the congruent demonstrative and gaze condition there was no significant difference between the single versus congruent proximal, *OR*=1.33, *SE*=0.20, z=-1.05, p=1.00, or distal cues, *OR*=1.69, *SE*=0.47, z=1.90, p=.86. Therefore, demonstratives alone were enough in guiding attention. However, when demonstrative and gaze were incongruent, there was a significant effect of (incongruent) gaze when the demonstrative was distal compared to both the single demonstrative condition, *OR*=2.67, *SE*=0.66, z=3.98, p=.001, and the congruent demonstrative plus gaze condition, *OR*=3.55, *SE*=0.81, z=3.98, p<.001. When the demonstrative was proximal, however, incongruent gaze did not affect the choice of referent compared to either the single demonstrative condition, *OR*=1.12, *SE*=0.29, z=.43, p=1.00, and the congruent demonstrative plus gaze condition, *OR*=1.52, *SE*=0.33, z=1.88, p=.894. In other words, gaze only impacted referent choice when the demonstrative was distal.

Figure 2.13: Interaction of demonstratives and eye gaze in Experiment 4 static trials. Error bars represent standard error.



For the dynamic stimuli, results also in the main replicated Experiment 3, with significant main effects as before of demonstrative, $\chi2$ (1) = 27.42, p <.001, pointing $\chi2$ (2) = 274.90, p <.001, and gaze, $\chi2$ (2) = 115.62, p <.001, and interactions between demonstrative and pointing, $\chi2$ (2) = 35.40, p <.001 and between pointing and gaze, $\chi2$ (4) = 124.98, p <.001, all yielding the same patterns as seen in Experiment 3 (see Figures 2.14 and 2.15).

Figure 2.14: Interaction of pointing and demonstratives in Experiment 4 for trials with videos as stimuli (dynamic stimuli).

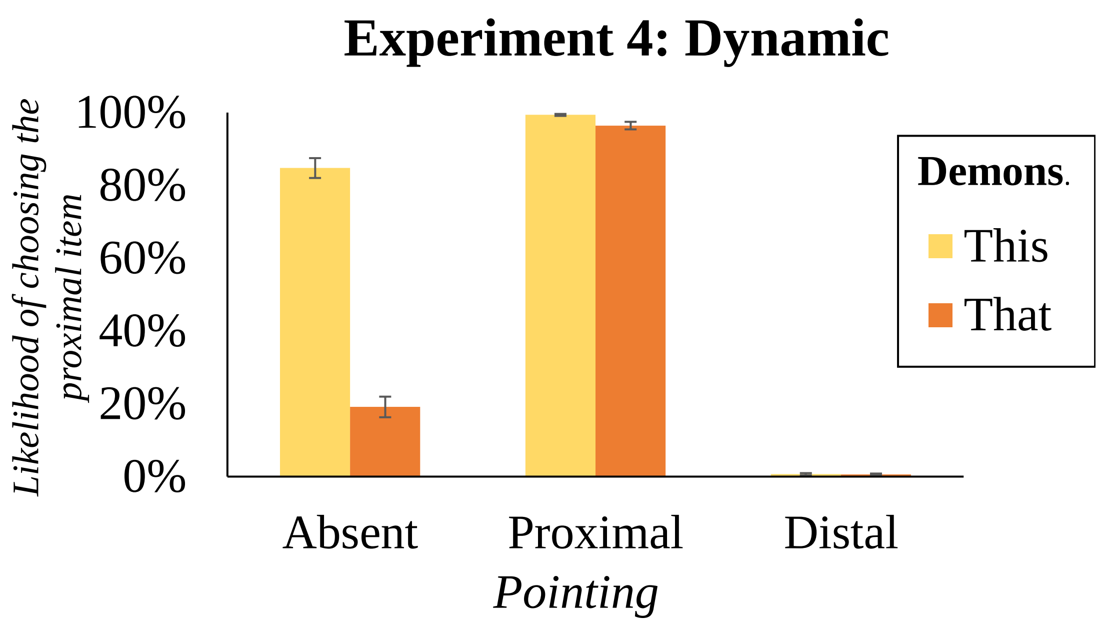
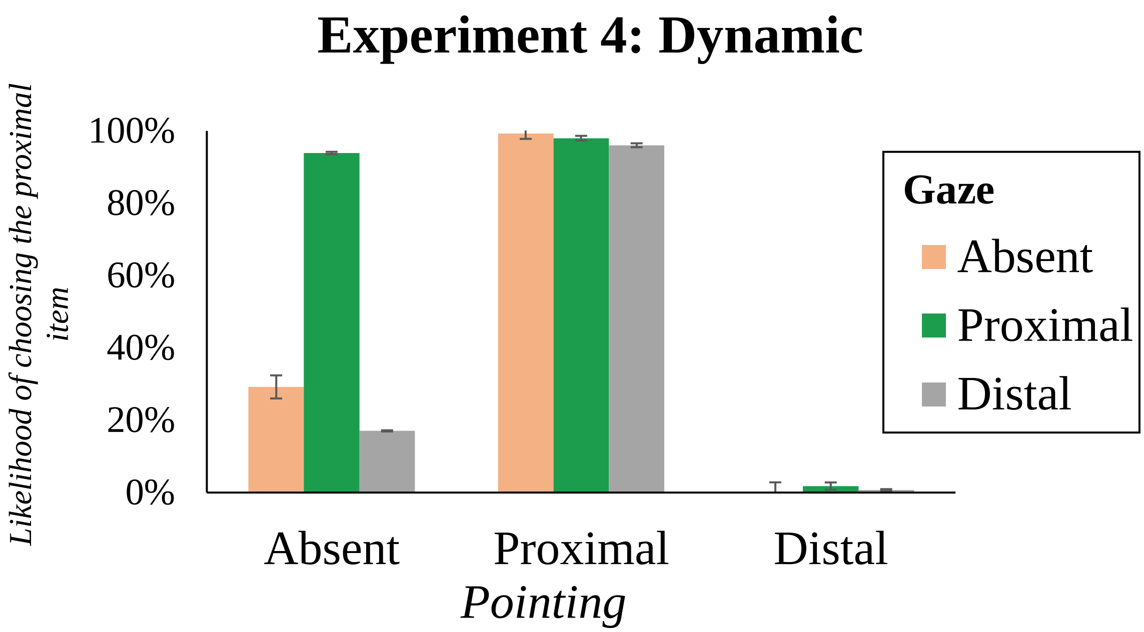
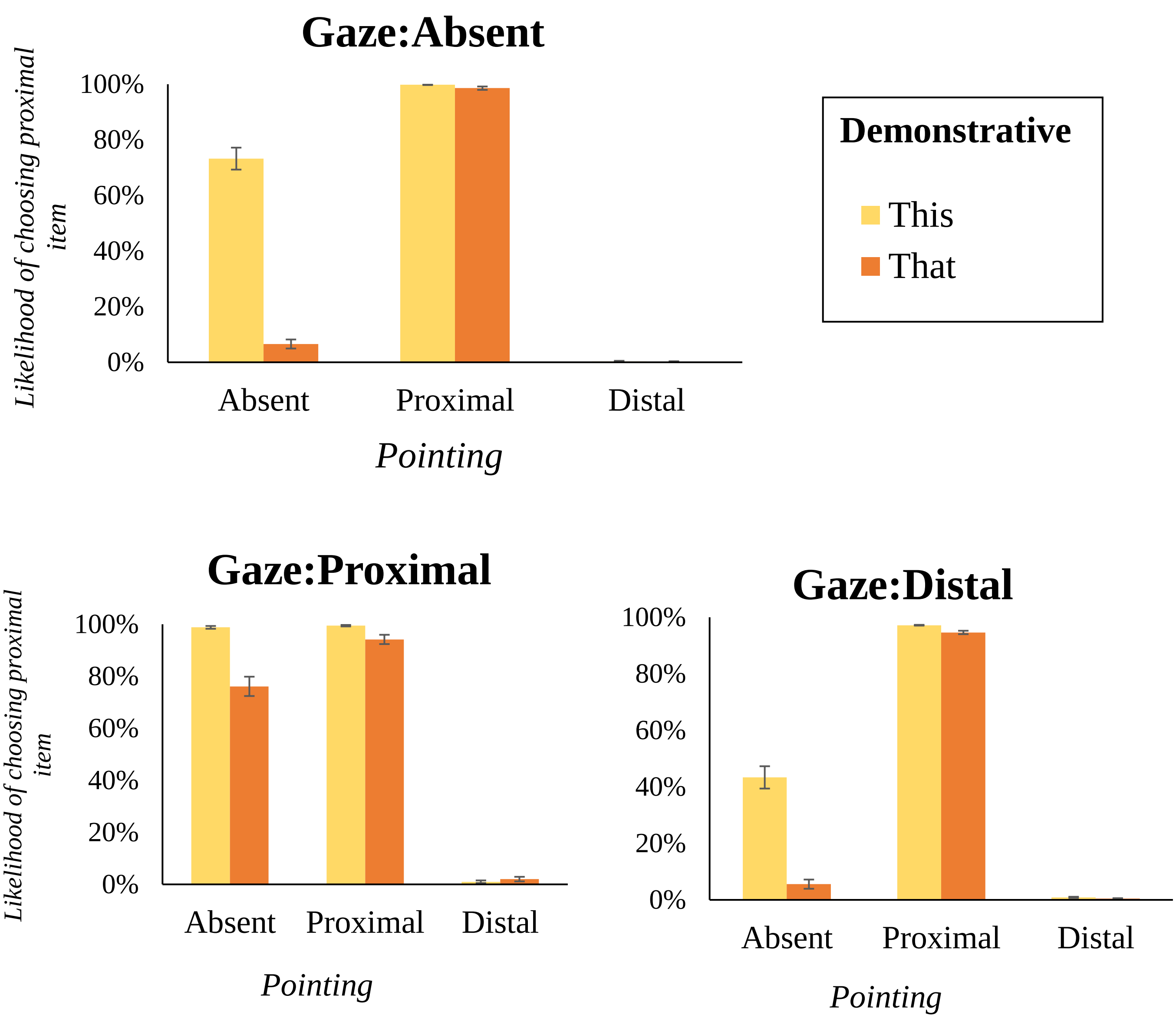


Figure 2.15: Interaction of pointing and gaze in Experiment 4 for trials with videos as stimuli (dynamic stimuli).



However, there was also a significant three-way between demonstrative, pointing and gaze, $\chi 2$ (4) = 10.69, p = .03 (Figure 2.16). In the absence of pointing, demonstrative and gaze affect referent choice for dynamic stimuli. For each level of gaze, *this* is associated with a greater choice of proximal referent compared to *that* (all contrasts p <.001). Moreover, congruence of gaze and demonstrative in the absence of pointing leads to a more decisive reference choice; proximal gaze and proximal demonstrative were associated with a greater selection of the proximal referent compared to the proximal demonstrative alone, *OR*=30.04, *SE*=13.39, z=7.63, p<.001. When cues (gaze and demonstrative) are incongruent, proximal gaze was the dominant cue even when the linguistic cue (demonstrative) was distal (*M*=.76, *SE*=.04) compared to trials when both cues were congruently targeting distal position (*M*=0.06, *SE*=.01), *OR*= 54.10, *SE*=16.20, z=13.32, p<.001. The three-way interaction also implied that, compared to trials without any non-verbal cue, having a combination of all three cues increased the strength of cueing to a target position (all contrasts $p<.001$. However, as long as the pointing arm was targeting the proximal item, participants would pick the proximal item, even when the gaze was distal and participants heard *that*, $OR= 1981.06$, $SE=1395.69$, $z=10.75$, $p<.001$ (compared to trials in which all cues target the distal position). The same trend is the case when pointing is distal; participants are cued by pointing even when gaze and demonstratives are proximal, $OR=2000$ $SE=.0004$, $z=12.87$, $p<.001$.

Figure 2.16: The three-way interaction of pointing, gaze and demonstratives.



The individual analyses above reveal differences in patterns of interactions for static and dynamic stimuli that drive the two- and three-way interactions of stimuli type with pointing, demonstrative, and gaze in the full model. Overall, the most noticeable difference between static and dynamic cues was the increased influence of (distal) gaze on object referent choice. When the agent was not pointing, gaze towards the distal position was a stronger cue in dynamic trials than in static trials, OR=3.57, SE=0.05, z=-7.22. The increased decisiveness of distal gaze was also observable at trials with an incongruent demonstrative. In dynamic trials, a combination of hearing the word *this* with distal gaze led to a decreased likelihood of choosing the proximal item, OR=5, SE=.04, z=-7.35 (Figure 2.17), compared to static trials. Overall, the effects of all cues are stronger in the dynamic compared to static trials. However, the data showed a similar pattern of effects for all three cues, both in Experiment 3 and Experiment 4 dynamic and static trials.

2.2.4.3 Discussion

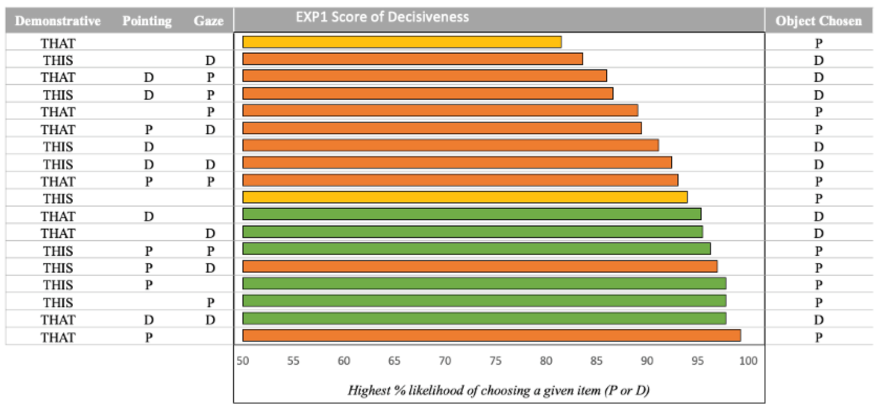
Results of Experiment 4 from all three analyses are broadly consistent with those of Experiment 3 regarding the significant effect of the three types of deictic cues individually and patterns of two-way interactions of pointing with gaze and with heard demonstrative. In addition, however, there are some specific differences as a function of stimuli type.

Results again support the view that demonstratives map onto perceptual space with the demonstrative *this* indicating object within reach/near the speaker and *that* indicating a farther referent (Anderson & Keenan, 1985; Caldano & Coventry, 2019; Coventry et al. 2008, 2014). Moreover, pointing and gaze were also associated with directing the participants to a particular referent, again consistent with the effectiveness of non-verbal cues as deictic pointers. Further replicating Experiment 3, pointing was the dominant cue, compared to both demonstratives and eye gaze. However, in Experiment 4, pointing and gaze together had an advantage over seeing one cue only, but only when targeting the distal position. This effect held even when data from static and dynamic stimuli were analysed individually.

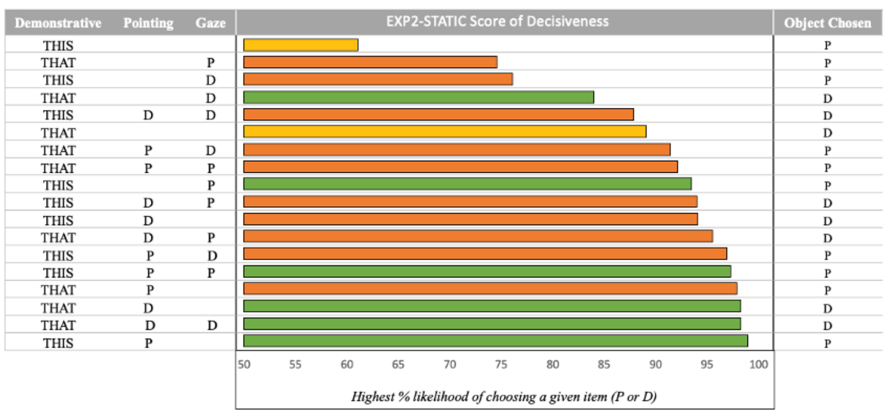
In the individual analyses separating static and dynamic cues, the pattern of interactions was overall consistent with the results of Experiment 3, but with two main differences. Firstly, in the static trials, there was a significant interaction between gaze and demonstrative, indicating that gaze affected referent choice when the distal demonstrative was uttered, but not when the demonstrative was proximal. The presence of this interaction in this Experiment (and the absence of it in Experiment 3) can potentially be explained by hearing the demonstratives rather than seeing it in a speech bubble, which is a more naturalistic way to perceive the verbal spatial reference. Second, in dynamic trials, there was a three-way interaction between pointing, eye gaze and demonstratives heard. This three-way interaction suggests that in the absence of pointing, eye gaze and demonstrative in combination are more decisive than hearing demonstratives alone. Furthermore, targeting with pointing, gaze and demonstratives towards the same position leads to a stronger effect, as opposed to hearing spatial demonstratives alone.

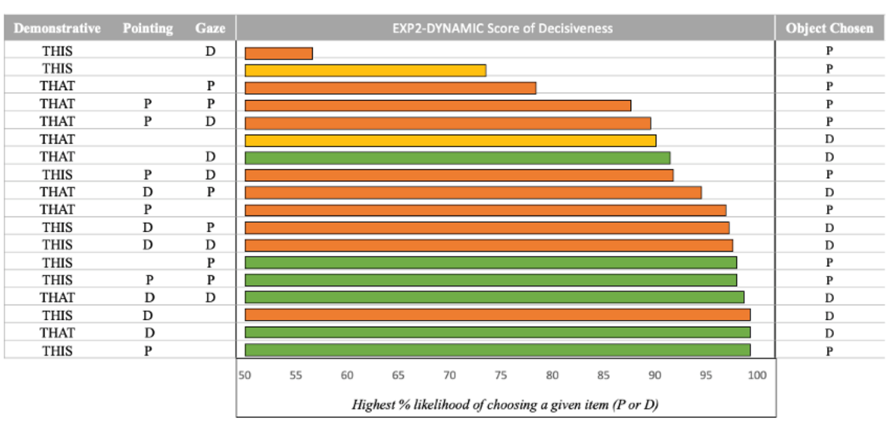
Overall, the differences between stimulus types suggest that seeing the action of referring which includes the trajectory of both eye gaze and pointing, and hearing spoken words uttered increase the decisiveness of the cues in combination (Figure 2.17). Static and dynamic trials mainly differed in terms of introducing the motion of referring as a factor. While the trends in choices based on two types of stimuli were mostly mirrored, with the introduction of dynamic trials, there is an increased decisiveness of all cues, in particular gaze and demonstrative. It is of note that the gaze detection task produced differences in the accuracy of the gaze detection task for static and dynamic stimuli. People are simply better at detecting gaze as a motion rather than fixated statically at the target position, which offers a possible reason for the increased effectiveness of gaze in particular in the dynamic scenes.

Figure 2.17: Ranked decisiveness score (raw means, not estimated marginal means) for each level of design across two experiments.



Labels (P and D) at the end of each bar represent which one of the potential two referents, proximal or distal, participants were most likely to choose at a given level of the design.
Orange=incongruent, yellow= single cue, green=congruent cues





2.2.5 General Discussion

In two studies our goal was to unpack the strength and relative effectiveness of deictic communication cues by manipulating the number and congruency of cues participants were exposed to on a given trial. In two experiments, we tested the potential advantages of seeing multiple cues over single cues on reference choice, and the relative strength of cues by pitting cues against each other using incongruent cues. Furthermore, we examined cues in the context of static versus dynamic deictic stimuli.

Figure 2.17 plots the decisiveness with which participants chose a referent, in terms of the overall percentage of trials where a referent was selected. As can be seen, all cues and combinations of cues lead to above chance levels of selection of a specific referent (proximal or distal), but a number of general patterns emerge. First, as the results of the experiments demonstrate, and as one can see in Figure 2.17, demonstratives on their own, while above chance cues to a specific referent, fair relatively worse than cases where there are multiple congruent cues. While supporting the mapping between demonstratives and perceptual space (Anderson & Keenan, 1985; Caldano & Coventry, 2019; Coventry et al., 2008, 2014), it is certainly not the case that *this* is associated with participants always selecting the proximal referent and *that* with choice of distal referent in the absence of other cues. A likely reason for demonstratives not being the most reliable source for referent choice is the inherent vagueness and flexibility in the use of demonstrative forms. Demonstratives not only refer to the objects in space but have other functions (Dixon, 2003). For example, it is well established that the use of demonstratives is affected by a range of factors, including the relative positions of speaker and hearer (e.g. Jungbluth, 2003; Coventry et al., 2008; Rubio-Fernandez, 2022), and the attention/gaze of the addressee (e.g. Küntay & Özyürek, 2006; Rubio-Fernandez, 2022). Furthermore, the choice of demonstratives is influenced by a range of non-spatial factors such as object visibility, familiarity and, ownership (Chandralal, 2010; Coventry, Griffiths & Hamilton, 2014; Skilton & Peeters, 2021). Given the multiple parameters associated with these terms, it is understandable in our data why on their own they are not perfect deictic pointers; they serve to communicate more than mere object reference.

Second, in general, one can see that congruent multiple cues (the green bars in Figure 2.17) are associated with more decisive referent choice compared to either demonstratives on their own (the yellow bars) or multiple incongruent cues (the orange bars), consistent with previous findings from Bangerter and Lowerse (2005). The increased decisiveness of multiple congruent cues over single or fewer cues may help to explain why multiple cues seem to occur in deictic communication (Carpenter et al., 1998; Talmy, 2018; Todisco et al., 2018), emerging synchronously relatively early in language development (Todisco et al., 2018).

Third, the pattern of results is also informative regarding the relative strengths of deictic cues. While demonstratives are relatively vague deictic pointers, the dominance of pointing over gaze and demonstratives was found consistently for both experiments and for both static and dynamic stimuli. In all cases when language and eye gaze were incongruent with pointing (see Figure 2.17), pointing most determined referent choice. While demonstratives are relatively vague deictic pointers, the dominance of pointing over gaze and demonstratives was found consistently for both experiments and for both static and dynamic stimuli. There are a number of possible reasons why this might be the case; perceptual salience, functional significance, and developmental precedence. We consider each of these in turn.

Compared to language and eye gaze, pointing may grab the most attention of viewers as a result of the perceptual salience of the action itself, both in terms the of size of a gesturing arm, and the availability of trajectory information (only Experiment 4). With static images, gaze occupies only a fraction of the image space compared to the pointing arm, making the pointing arm much easier to identify and process. Indeed, in our studies, we checked that participants could identify in the different gaze conditions which of the two potential referents on the table they thought the speaker was looking at. While performance was well above chance, it was nevertheless not perfect with static scenes, suggesting that gaze may be a less reliable visual cue than gesture. Moreover, with dynamic scenes, again the movement of the arm occupies more points changing in location, with the pointing moving nearer towards the referent than the gaze, which remains bound to the head. So, there is less projection required on behalf of participants from the speaker to the referent with pointing compared to eye gaze. The trajectory (pre-stroke, stroke, post-stroke hold and retreat) of the referential gesture overall is rich and elaborate in giving the listener cues for reference, making it a reliable and strong cue (Shattuck-Hufnagel et al., 2016; Kendon, 1980; McNeil, 1992; Norris, 2011). Overall, pointing improves the perceptual accuracy of locating a referent (Kowadlo, Ye & Zukerman, 2010). Langton et al. (1996; 2000) further argue that pointing (as well as head and gaze orientation) is processed automatically.

Another possibility is that participants may be cued to attend to gesture given that gesture functionally precedes both language and gaze, making gestural information available earlier than other modalities. People tend to begin gesturing milliseconds before the speech (Church, Kelly & Holcombe, 2014) and hold the gesture after the speech, making it available for longer than language. There is then an argument that gesture may grab the attention of the hearer/viewer first and hold it longest, giving more time for participants to locate the intended referent.

Yet another possibility is that pointing occurs early in development and may therefore take precedence on developmental grounds. Pointing is an important landmark/turning point in the development of language. There is a large body of research suggesting that deictic gestures are a universally prelinguistic behaviour closely tied to the emergence of the first words of children (Carpenter, Nagell & Tomassello, 1998; Iverson & Goldin-Meadow, 2005). Children at around 9 to 12 months of age engage in gesture use while communicating (Liszkowski, 2010), for example, with their caregivers, before they start talking. Furthermore, the use of gesture precedes, predicts and overall, closely relates to (efficiency in) vocabulary learning and the emergence of language use (Carpenter, Nagell & Tomasello, 1998; Iverson & Goldin-Meadow, 2005; Liszkowski, 2010). While it is unclear why pointing is universally present in infants prelinguistically, the 'gesture-first' hypothesis maintains that gesture preceded language in the development of communication (e.g., Arbib, 2012; Corballis, 2010; Tomasello, 2008), and as such the precedence of gesture developmentally could link to its precedence diachronically.

Taken together, there are compelling reasons why pointing in our data is the most persuasive cue to deictic reference, especially when cues are incongruent. However, it is also the case that, in general, multiple (congruent) cues lead to more decisive referent choices than fewer cues. Although there is evidence for the co-occurrence of multiple cues in the early use of deictic communication (e.g., Todisco et al., 2020), and in some languages, it is obligatory to point or gaze when using spatial demonstratives, there is also recent evidence suggesting that pointing gestures do not always accompany demonstratives. For example, Bangerter (2004) found that gesturing was more likely to occur when a referent was in close proximity to a speaker (see also Cooperrider, 2016), but pointing was less frequent when using distal demonstratives to refer to far locations (see also Piwek et al., 2008). If pointing improves the localisation of a referent, as was the case in our studies, one can ask why people do not always point when using demonstratives. One key difference between these studies and ours is that the further object in our studies was still not especially distal to the speaker, whereas the Bangerter (2004) studies involved referents that were much further away from participants. Diessel and Coventry (2021) suggest that the use of pointing in combination with distal referents may be predicted by the extent to which pointing can unambiguously ‘pick out’ the intended referent (Lücking, Pfeiffer & Rieser, 2015). Accordingly, one can predict that rerunning our studies with much more distal references would correspondingly reduce the dominance of pointing as a decisive cue to the speaker’s intended referent.

The use of multiple deictic cues also provides a means for a speaker to communicate much more information about an object being referred to than mere object location. Indeed, a mismatch between deictic cues affords communication of information about the connection and attitude of a speaker towards the object referred to. For instance, the use of *that* when pointing and looking at a proximal object, while drawing the attention of the hearer to the object, may also provide information that the object is not liked by the speaker, or that the object is not a main attentional focus for the speaker and may be less likely to occur in future reference. The multifaceted nature of deictic communication places it at the junction between spatial communication and social communication, making it informationally rich and a natural bridge between spatial understanding and theory of mind (Rubio-Fernandez, 2022).

In summary, results from theseexperiments suggest that pointing is the most dominant cue as a determinant of referent choice and that demonstratives on their own are rather imperfect cues to object location. When pointing is incongruent with either gaze or demonstratives, pointing is the strongest predictor of choice of referent participants think the speaker is referring to. Proximal cues (including the word *this*) led to a higher likelihood of picking the proximal referent individually, and the effect is strengthened when all three cues are present. The effect of demonstratives is also strengthened with the addition of a non-verbal cue. Future studies would do well to examine online measures of attention to cues, for example with the use of eye tracking, to examine the primacy of focus on non-linguistic cues to observe the trajectory of visual scanning when perceiving deictic cues. Attention to response time can also be informative regarding whether incongruence in any one specific or a combination of deictic cues would lead to longer response time, implying difficulty in processing deictic information.

2.3 Concluding remarks

All four experiments suggested pointing as the dominant cue. Unimodally, language implied a stronger effect as the stimuli became progressively more naturalistic (similar change in effect was the case for eye gaze too). The addition of at least one non-verbal modality influenced the processing of linguistic cues to reference, both congruently and incongruently. In the final experiment, we can indeed say that two cues are better than one, which is in line with previous work on non-deictic gestures.

Chapter 3 Multisensory Deictic Communication

As we have already established in the previous chapters, deictic communication is closely associated with perceived space. Objects within reach are more likely to be labelled with a proximal demonstrative (peripersonal space) and outside reach (extrapersonal space) with a distal one. While peripersonal space is the only region of space which benefits from multisensory input in its entirety, most research on deictic communication exclusively focused on vision. In the following section, we report an experiment which probed the nature of stimuli within peripersonal space with a specific contrast of visual and haptic spatial perception. The experiment is a novel haptic adaptation of Memory Game Paradigm.

3.1 Experiment 5

3.1.1 Introduction

Spatial demonstratives (e.g., *this*/*that/here/there* in English) are among the most important terms in all languages. They are universal (Deutscher, 2005; Diessel, 1999; Diessel & Coventry, 2020) and they are among the first words in a child's lexicon (Clark, 1978; Diessel & Monakhov, 2022). Furthermore, the use of demonstratives is often multimodal. Demonstratives are the terms in language most associated with pointing and eye gaze (Enfield, 2003; Stukenbrock, 2015; Levinson, 2018); in some languages, it is obligatory to point or gaze when using specific demonstrative forms (e.g. *Yélî Dnye*, Levinson, 2018; *Gomai*, Hellwig, 2003; *Kilivila*, Senft, 2004; *Yucatec*, Bohnemeyer, 2018; *Warao*, Herrmann, 2018; *Tiriyó*, Meira, 2018). From an early age, children use demonstratives, eye gaze and pointing in synchrony when directing a conspecific to an object location in the environment (Todisco, Guijarro-Fuentes, Collier & Coventry, 2020). Here we explore how demonstratives in English are used as a function of the senses involved in perceiving the spatial world. The very nature of different senses (vision versus touch) leads to contrasting predictions regarding how demonstratives might be produced when exploring the spatial world in different ways.

While it has been established that choice of demonstratives is affected by a wide range of parameters, including the nature of the objects being referred to (e.g. object visibility/ownership/graspability; Coventry et al., 2014; Rocca et al., 2019a), the attention of speaker and addressee (Küntay & Özyürek, 2006; Rubio-Fernandez, 2022), the relative positions of speaker and addressee and their interaction (Coventry et al. 2008; Jungbluth, 2003; Rubio-Fernandez, 2022; Rocca et al.,2019b), and the environmental context (e.g. elevation; Forker, 2020), the idea that demonstratives map onto space has a long history. In linguistic typology, it has often been assumed that demonstratives map onto distance, with a term for reference to a position near the speaker (e.g., *this* cup) and a term for a position further from the speaker (e.g., *that* spoon) (Anderson & Keenan, 1985; Bühler, 1934; Diessel, 2014). However, there is much recent evidence in English that *this/that* is associated with the more precise reachable/non-reachable space distinction (Coventry et al., 2008, 2014; Caldano & Coventry, 2018; see also Kemmerer, 1999, who originally suggested such a mapping but rejected it).As described in Chapter 1, using a 'memory game' paradigm, where participants point and describe object locations, it has been shown that *this* is preferred when the object being referred to is reachable by the speaker, with a gradual shift from *this* to *that* as the object moves outside of reach to non-reachable space (Coventry et al., 2008). Further investigating the reachability hypothesis of demonstrative choices, Caldano and Coventry (2018) asked participants to point at objects in locations reachable with one hand but not the other; *this* was used more when the pointing hand could reach the object than when the pointing hand could not. Evidence for a mapping between reachable/non-reachable space and spatial demonstrative use has now been established across a wide range of languages including Estonian and Võro (Reile et al., 2020), Ticuna, Dutch (Skilton & Peeters, 2021) and Spanish (Coventry et al., 2008), and recently a major study has argued that such a distinction may be a semantic universal, with evidence of a mapping between reachable/non-reachable space and proximal and distal terms across 29 diverse languages (Coventry et al., in press).

The behavioural results on demonstrative use are consistent with a mapping between demonstratives and the perceptual distinction between peripersonal space and extrapersonal space, associated with different underlying brain systems (Berti & Rizzolatti, 2002; Bufacchi & Ianetti, 2018; Legrand, Brozzoli, Rosetti & Farné, 2007; Làdavas, 2002; Gallivan, McLean & Culham, 2011; Makin, Holmes & Zohary, 2007). PPS is "a network of body-part-centred representations responsible for the coordination of actions toward, and avoidance of, objects and other living entities" (Hunley & Lourenco, 2018, p14; see also di Pellegrino & Làdavas, 2015).As briefly mentioned at the general introduction, the boundary of PPS space is plastic; as reach is extended through tool use, so, too, is the use of *this* to the region reachable with the tool (beyond the hand; Coventry et al., 2008). Importantly, PPS is multisensory; evidence from neurophysiology and neuroimaging studies shows that PPS representation is implemented by specific neuronal populations, which selectively integrate tactile stimuli with visual or auditory cues related to external objects as they approach the body (Làdavas & Serino, 2008; Macluso & Maravita, 2010; Cléry et al., 2015, 2017; see Serino 2019 for a review). While PPS is multisensory, to date there are no studies examining demonstrative use across different sensory conditions (e.g., through touch alone). However, if demonstratives *do* map onto perceptual space, then experiencing the world through another modality, such as the haptic modality alone (active touch with reach to grasp without vision) may still be associated with spatial descriptions illustrating a PPS-EPS distinction.

However, there are equally compelling reasons to argue that differences between perceptual modalities, by their very nature, may lead to differences in spatial demonstrative use. While the visual modality allows access to objects remotely providing constant information regarding distance and relative object distance of the intended referent with respect to the body and environment, the tactile modality requires that objects are contacted to know where they are, with distance gauged through locomotion back to the point of description. There is evidence (from studies with vision) that contacting an object itself affects demonstrative use. Coventry et al. (2008; 2014) had participants (or someone else) place objects on a table at different locations prior to spatial description, finding an increased use of *this* when participants placed the object. This 'who places' effect is explained by Coventry et al. (2008) in terms of the presence of the object in the participants' PPS prior to object placement. As the haptic modality requires processing the object in PPS, one can predict that the haptic modality may therefore lead to increased use of *this*. This effect may be further exacerbated by the dominant role that vision plays as a support for other sensory modalities (Aggius-Vella et al., 2018; Klatzky, Lederman & Matula, 1993; Lederman et al., 1990; Newport et al., 2002; Pasqualotto et al., 2013), even when the information is unrelated (Zuidhoek et al., 2003; Postma et al., 2008). The absence of vision and visual information may lead to slight distortions in spatial perception, processing and cognition across different senses (Cattaneo et al., 2011; Kappers, 2002; Kolarik et al., 2017; Lederman et al., 1987), indicating increased reliance on the initial memory of contact with the object in PPS.

In the experiment below, we examine for the first time how people use demonstratives under three different sensory conditions - vision plus haptic, vision only, haptic only – to examine whether the mapping between PPS/EPS persists in structuring demonstrative across different senses, or alternatively whether the absence of vision and nature of haptic contact with objects may lead to the dominant use of the proximal term.

3.1.2 Methodology

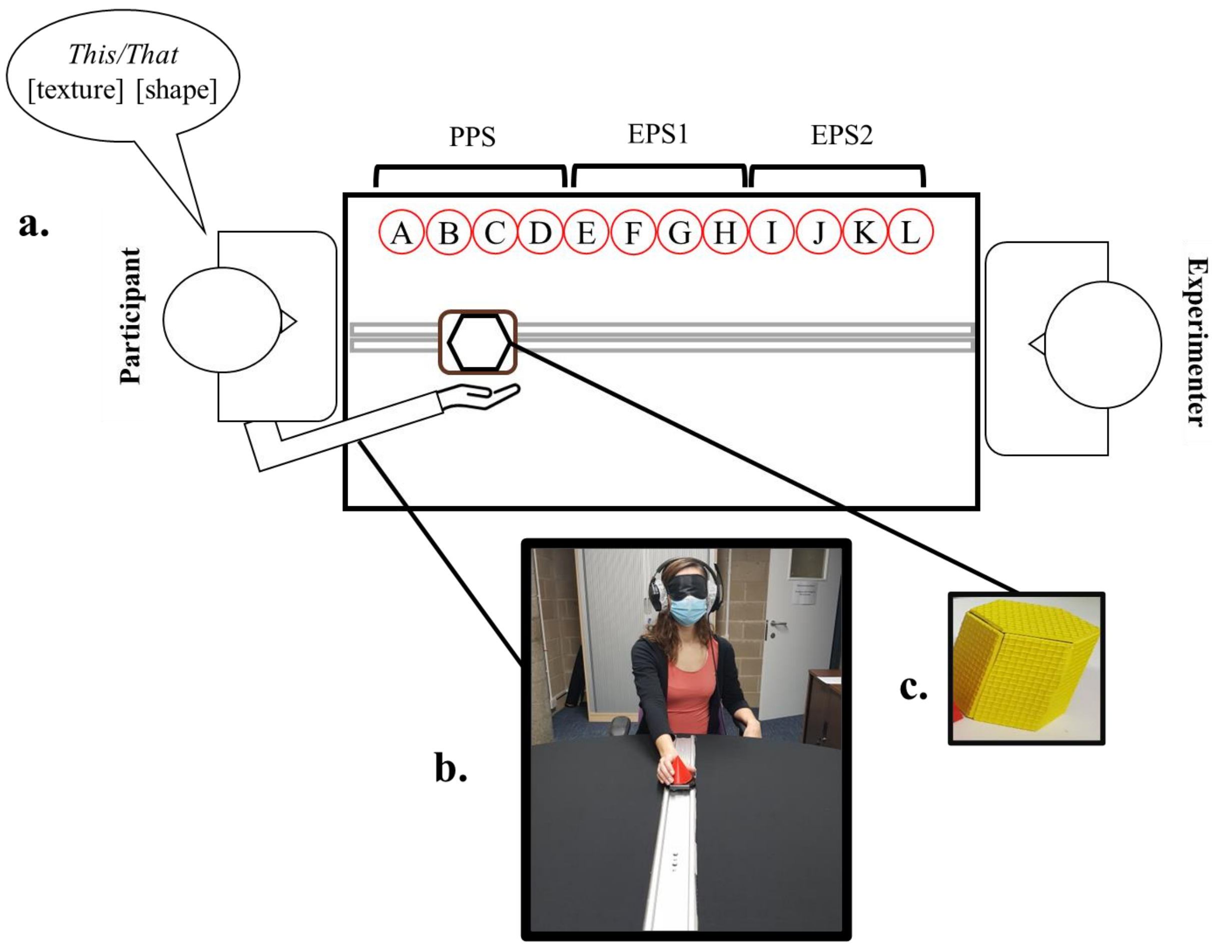
Participants: Power analysis was carried out using G*Power (version 3.1) (Faul, Erdfelder, Buchner & Lang, 2009) to determine sample size, with effect size based on a prior pilot study (Arikan, 2018). To detect a large effect (f = 0.35) at a significance criterion of α = .05, a sample size of 26 participants per condition for this experiment was indicated. Eighty-one monolingual English-speaking participants (M_{age}=25, Female=62) took part for payment or course credit with 27 participants in each of the three conditions (vision only condition, haptic only condition, vision plus haptic condition). Each participant was screened (self-report) for health conditions which might affect their haptic perception, such as diabetes mellitus or anxiety disorders such as nyctophobia which could be provoked by blindfold use. The experiment took around 40 minutes to complete.

Design, Stimuli and Procedure: The experiment adopted the 'memory game' paradigm (Caldano & Coventry, 2018; Coventry et al., 2008, 2014; Gudde, Griffiths & Coventry, 2018) described at the general introduction. In the original paradigm, the referents were disks with images, and were presented at different distances in front of participants (see Figure 1.1). Similarly, participants are informed that they are taking part in a study exploring the effects of language on memory for object location and that they are in the language condition. (Memory is a cover so that language is incidental to the study from the participants' perspective; see Gudde, Griffiths & Coventry, 2018 for discussion.) Following the placement of an object at different locations by the experimenter (effectively the 'addressee'), participants (the 'speakers') verbally refer to each object using three words - a demonstrative (*this* or *that* in English), an adjective (e.g., colour) and the object name (shape), e.g., *this red star*. Participants are informed that they need to stick to using only three words so that the amount of language used is the same for all people in the language condition. Memory probe trials at various points during the experimenter keep the memory cover active throughout the task.

The current study adapted the memory game paradigm for use across different sensory conditions (see Figure 3.1). Participants were informed prior to consent that they were taking part in a study exploring the effects of language on memory for object location under different sensory conditions and that they would look, touch or look at and touch objects to remember their positions. Participants were randomly allocated to one of the three conditions - haptic only, vision only (looking at the objects), and a combination of haptic and visual inputs. Participants in the haptic condition wore a blindfold throughout the task. The participants in the visual and combination conditions wore the blindfold only when the experimenter was placing the object at the target location (to control for object exposure across conditions). Three-dimensional objects of the same size but with different shapes and textures were used as the referents in order to give objects haptically distinguishable properties that were also visually distinguishable (e.g. rubber sphere, rough cylinder, etc.). The use of three-dimensional objects is one of the key differences from the original memory game paradigm.

Figure 3.1: Schematic representation of experimental set-up.

a.) The objects were placed on a 150cm rail, one object at a time. The 12 object positions used are represented by letters A to L. The first four objects are within arm's reach, therefore PPS, while further distances were not (all in EPS.) An image of a participant doing the task. c.) An example object used during the experiment.



On each (experimental) trial, there were two phases - learning and labelling - each signalled with an auditory cue. During the learning phase (cued with a low pitch sound), participants located the objects and tried to memorize the object's position on the rail. In the haptic condition, participants reached towards the object to locate it. Note that for objects in EPS (Figure 3.1), this involved stretching (for EPS region 1) and standing and stretching (for EPS region 2). In the visual condition, participants saw the object without reaching/touching, remaining in the same seated position throughout. In the combination condition, they used both sensory modalities. After ten seconds, the end of the learning phase and the beginning of the labelling phase was cued with a higher pitch sound. During the labelling phase, participants used three words to refer to the object. After every three experimental trials, there was a memory probe trial to maintain the memory cover for the experiment, cued with a third distinct sound. For memory trials, participants were instructed to move one of the objects from the last three trials to where they thought it had been located by pushing it up the rail.

Participants were instructed that they should point towards the object (arm outstretched and pointing with index finger), labelling the object position using a combination of a demonstrative word (either *this* or *that*), an adjective (the texture of the object) and noun (the shape of the object), e.g., *this/that rough pyramid*. Moreover, they were also asked to try to use both demonstratives during the task and were reminded to do so after the practice trials if they only used a single demonstrative during the practice trials. Instructions never referred to the use of a specific demonstrative at a specific location to ensure that participants were not primed to use demonstratives for specific locations. The objects were placed by the experimenter at specific locations one object at a time on a 150 cm-long rail placed on a table (Figure 3.1). There were 12 equidistant locations, with 4 locations each falling into three regions on the sagittal axis: within

PPS (reachable with ease while sitting), or two regions in EPS (EPS1: reachable with outstretched arm while sitting; EPS2: reachable only by standing up and stretching forward with effort). The presentation order of the object positions and the type of object allocated to a specific position were randomized across trials. Each participant took part in 48 trials across four blocks, with the 12 locations randomised in each block.

After obtaining informed consent, a series of preliminary physical measures were taken from the participants including arm length, height and three reaching distances consisting of their arm resting on the table, leaning forward on the table with a stretched arm while seated and leaning forward on the table with a stretched arm while standing. These measures were taken to calculate object positions for each participant so that distances A-D (Figure 3.1), for example, were within PPS for all participants. The participant-specific calculation for referent position was another key difference from the original paradigm.

Participants then entered the experiment room with the blindfold on and the experimenter helped participants to their seat. Before the experiment started, participants were familiarized with the stimuli to check they were able to name them correctly. The experimenter showed/gave each object to participants and labelled them e.g., rough cylinder, without using any demonstrative. As part of familiarisation, participants in the haptic and combination groups also touched the rail in front of them and were guided to the farthest possible position they would have to reach. Familiarization was followed by three practice trials. During the familiarisation, participants also heard the auditory cues and learnt their meanings, cueing phases of experimental trials and the memory probe trials. Participants were reminded of the instructions after familiarisation and again after the practice trials. If participants failed to use a demonstrative for 3 consecutive experimental trials, they were reminded of the rules once more, and if a participant failed to use a demonstrative after being corrected and therefore misunderstood the task, the data collection was stopped (also see Gudde, Griffiths & Coventry, 2018). Throughout the experiment, participants wore noise-blocking headphones (with white noise) so that they remained focused on the task at hand (described below) and to make sure there were no auditory cues to object location as the object was moved/placed. They heard the auditory cues through the noise-blocking headphones.

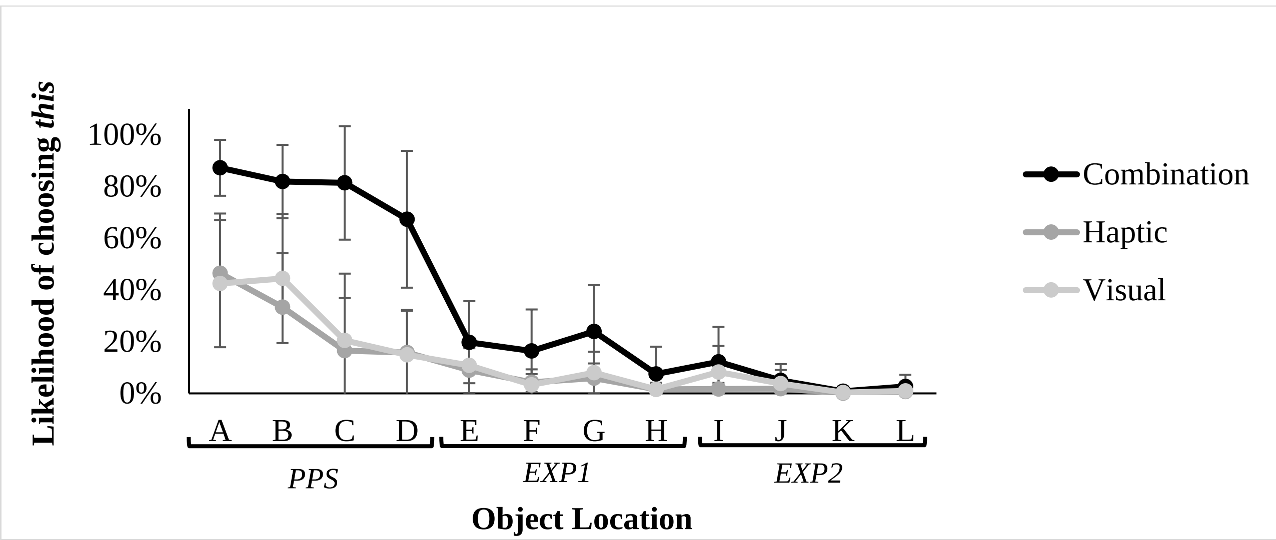
At the end of the experiment, participants were fully debriefed (verbally and written) and were asked if they guessed the aim of the study (participants who realised that the study was about demonstrative choice were excluded, in accordance with the memory game protocol). Two of the participants guessed the aim of the study and we recruited two more participants to meet the sample size criteria (so that there were at least 26 participants in each condition).

3.1.3 Results

The experimental data were analysed using generalized linear mixed model using Jamovi 2.2.5, (Gallucci, 2019; R Core Team, 2021) with binomial distribution and logit link function due to the binomial nature of the outcome (this=1, that=0). The word choice was set as the dependent variable while object position (A-L) and sensory modality condition (haptic only, visual only, combination) were allocated as fixed effect factors. Factors were coded with Helmert contrast. We included random coefficients with participants as intercepts, we also included the main effects of predictors as random slopes (as long as the model converged model fit was not singular). The random effects structure was formed considering the convergence and theoretical relevance (Barr, Levy, Scheepers & Tily, 2013). Final random components included Intercept | Participant and Participant | Condition. The main effect and interaction of the predictors were reported from fixed effects omnibus tests. The fixed effect coefficients were assessed using Wald’s Z statistics (Table 7.7).

Significant effects and interactions were further investigated using posthoc analysis with Bonferroni correction. The model was estimated using log likelihood (odds ratio, OR) and OR is reported for model estimates. We report inverse odds for levels of analysis with OR less than 1 (1/OR). Graphs and directional data reported are obtained from the estimated marginal means (EMM).

Figure 3.2: Graph representing the estimated marginal mean (in percentage) use of the word *this* for the 12 object positions by condition. Error bars represent standard error.



The model testing main effects and interactions (Conditional R^2=.88) of the sensory domain used for spatial perception and object position, yielded a significant main effect of object position, $\chi2$ (11) = 367.84, *p* <.001 (see Figure 3.3). The type of sensory input used to perceive object location was not a significant factor, $\chi2$ (2) = 2.21, *p* =.33, and the interaction of object location and sensory domain was also not significant, $\chi2$ (22) = 15.15, *p* =.86 (Figure 3.2).

Post-hoc contrasts (with Bonferroni correction) for the effect of distance showed that responses to the object locations within PPS, from locations A to D, did not significantly differ from each other in the likelihood of picking *this* (all contrasts $p>.05$). However, the first four positions significantly differed from any position outside of PPS (all $p < .001$). For the rest of the position contrasts outside PPS, there was only a significant difference in the likelihood of choosing *this* between Position E (the first object position in EPS1) with Position K, and marginally with Position L (the last two object positions in EPS2), OR=2.86 SE=.83, z=3.63, p=.019 and OR=2.62 SE=.75, z=3.39, p=.047 respectively. Both of these contrasts suggest that participants were more likely to choose the word *this* at position E, the nearer one of the three.

3.1.4 Discussion

We set out to examine how people use demonstratives as a function of the experience of space through different sensory modalities: vision, only, haptic only, vision plus haptic modalities. This was motivated by two contrasting hypotheses. On the one hand, one can argue that experience of PPS is multisensory, and within PPS there would be no difference among the senses regarding perceived distance. Therefore, participants should show similarities in demonstrative choices across conditions. On the other hand, different senses change how the spatial world is explored, with potential consequences for demonstrative choices. In particular, grasping objects in the haptic modality involves contacting objects in PPS objects and gauging distance back to the point of description, leading to a predicted increase in the use of *this* (consistent with 'who places' effects found in Coventry et al., 2008, 2014). We further prospered that there could be an even stronger increased use of *this* in the haptic-only condition compared to the haptic plus vision condition given the importance of vision as means of calibration for other senses (e.g. Aggius-Vella et al., 2018).

The results show no differences in demonstrative use as a function of sensory modality. Across all conditions (vision, only, haptic only, vision plus tactile) there were strong effects of PPS versus EPS on demonstrative use, with a graded fall-off in the choice of *this* to describe object location. This is not only consistent with previous studies showing the graded nature of PPS (e.g., di Pellegrino & Làdavas, 2015), but also with past studies showing demonstrative use as a function of distance and object reachability in tabletop space (Coventry et al., 2008, 2014, in press; Caldano & Coventry, 2018). This pattern of results for the haptic-only (blindfold condition) shows that vision is not necessary for a referential communication task for demonstrative distinctions based on perceptual space to occur, consistent with the multisensory nature of PPS.

One can ask why the present results in the haptic-only and haptic plus vision conditions do not show a similar increase in the use of *this* that was found in previous studies using the memory game paradigm (Coventry et al., 2008, 2014). Coventry et al. (2008) had participants (or the experimenter/addressee) place objects prior to the description of the object position. In doing so, the participant moved the object by picking it up and physically moving through space to place it before returning to his/her seat. In comparison, participants in both the haptic and haptic plus vision conditions in the present experiment merely briefly grasped the objects once placed by the experimenter, and only moved the object as part of the memory cover task. While Coventry et al. (2014) argue that the who places effect may be a result of the presence of the object in the PPS of the participants during placement, the present results discount that explanation. Rather, the 'who places' effect in previous studies may be a consequence of agency rather than merely the presence of the object in the PPS of the participant.

The absence of the effect of sensory experience on spatial language suggests that senses operate as independent modalities for spatial perception (Ottink et al., 2021). Although a body of past studies has proposed that vision may be essential for accurate spatial perception as a means of calibrating other modalities (Aggius-Vella et al., 2018; Klatzky, Lederman & Matula, 1993; Lederman et al., 1990; Newport et al., 2002; Pasqualotto et al., 2013), in this task requiring haptics and proprioception the absence of vision did not have a significant effect on word choices. Such a result is consistent with perceptual work exploring PPS with blind individuals. Ricciardi et al. (2017) compared blind and sighted participants on pantomimed reach-to-grasp actions at different positions in space, concluding that lack of visual experience does not influence PPS representation. This begs the question of whether the similarities and differences in the spatial representation of blind adults are mirrored in their use of deictic expressions. There is some evidence that in general spatial language in blind individuals is not different from the sighted (Andersen, Dunlea & Kekelis, 1984; Rosel et al., 2005). Blind children start to use spatial terms as early as sighted children, although the frequency of demonstratives is lower in blind children (Perez-Pereira, 1999). Gesturing, which co-occurs with demonstratives, also occurs with blind individuals although the frequency of gesturing seems to differ (dependent on task) compared to sighted individuals (Iverson & Goldin-Meadow, 1998;2001; Mamus et al., 2023). Mamus et al. (2023) recently suggested when describing motion events auditorily, blind individuals use pointing gestures more often than sighted individuals to indicate positions of landmarks. Future studies would do well to consider investigating the nature of deictic communication comparing sighted and (late and early) blind individuals in order to explore the use of deictic expressions across sensory modalities and across types of spatial localisation (e.g., haptic versus auditory).

In summary, the present work has found that the mapping between the use of the proximal demonstrative *this* and PPS and the distal demonstrative *that* and EPS seems to be robust irrespective of the sense or combination of senses used to perceive object location in space. Such a finding bolsters the view that that mapping between reachable and non-reachable space may be of fundamental importance to structuring of the English demonstrative system.

Chapter 4 Implications of visual loss for deictic communication

In previous chapter, I presented the results of a study looking at the implications of the multisensory nature of perception on the mapping of language to space. Results suggested no significant difference amongst senses, showing a robust effect of reachability across senses. This raises the question of whether we would observe such similarities for a population that always relies on non-visual means of perception, blind and visually impaired individuals. In this chapter, I briefly review literature on influence of blindness on spatial perception and communication. I then report Experiment 6, a pilot study where with a group of visually impaired individuals who took part in the haptic adaptation of the memory game paradigm described in the last chapter.

4.1 Spatial perception in blindness

"Blindness takes away one's territorial rights. One loses territory. The span of attention, of knowledge, retracts so that one lives in a little world. ... Only the area which can be touched with the body or tapped with the stick becomes a space in which one can live. The rest is unknown." John M. Hull, *Notes on Blindness*.

In his book *Notes on Blindness*, Hull used these words to describe his gradually developing blindness. From using other bodily senses to understand the world, to forgetting what his wife and children's faces looked like, we can understand the alternative ways in which reality is shaped, solely based on senses used. In this quote, he emphasizes how loss of sight leads to a shrinkage of space and territory of belonging, a drastic change in the world around him. Previous work on blindness has established that the absence of vision leads to a significant change in the way one interacts with the space around, influencing communication of contextual information (Saerberg, 2010). However, little is known about spatial communication in blind people, with surprisingly limited research in this area (as we shall see below). The changes in perception associated with blindness naturally leads to the question of implications of the absence of vision on the language of space.

The population of individuals with visual impairment and blindness is growing with leading causes including cataracts, macular degeneration, glaucoma, and diabetic retinopathy. According to the National Eye Institute, the number of people with visual impairment or blindness in the United States is expected to double to more than 8 million by 2050 while this number will double to four million by 2050 in the UK (RNIB). The population with blindness is increasing proportionately to the aging population.

Loss of sight influences lives of the individuals in number of social and psychological ways in their daily lives (Kemp,1981) including working conditions (Kunz et al., 2014; Steverson & Crudden, 2023), educational settings, recreational activities (Jones, Murray & Gomes, 2022) such as visiting museums (Hutchinson et al., 2020; Schales & Chastain, 2023) and participating in sports (Mannella, 2021). Candin (2003;2004) argues that blind and visually impaired individuals experience "an ocularcentric bias" in a sighted world where the general public and regulations are vision-based and restricting for people without sight. However, there are new technologies to help blind and visually impaired individuals to experience an inclusive world (e.g., Khan et al., 2020; Tuttle & Carter, 2022).

There are also psychological effects of visual loss and blindness (Thurston et al., 2010). People with visual loss and blindness tend to be vulnerable to mental health problems such as depression, anxiety (Augestad, 2017; van Munster et al., 2021; Osaba et al., 2019) and loneliness (Veerman et al.,2019). These mental effects can be alleviated with, for example, developing a sense of group identity and group work (Mannella, 2021; Zapata, 2022). Visual impairment and blindness also affect the people around the visually impaired individuals, for example their families (Chen & Groves, 2023; Tavakol et al.,2008).

Our early sensory experiences with the outside world shape the ways we process information (Williams, Huang & Bargh, 2009), form mental representations of our surroundings (Marschark & Mayer, 1998) and organization of memory networks (Nelson, 2001; Cardin et al., 2018). In the absence of vision, blind and visually impaired individuals rely on their sense of touch and hearing to navigate around and perceive distance. Although vision is argued to be the dominant sensory modality of spatial perception, blind people still develop spatial skills comparable to sighted due to cortical plasticity (Bedny & Saxe, 2012; Bonino et al., 2008; Cattaneo et al., 2011a; Jones, 1975). Non-visual senses are argued to compensate (Cattaneo & Vecchi 2011) and substitute visual information in forming spatial relations (Chebat et al., 2018). However, blind individuals might still have difficulties in certain spatial tasks such as serial order in spatial processing (Bottini, Mattioni & Collignon, 2016) and shifting of reference frames (Martolini et al., 2021).

The literature on the spatial capacities of blind individuals is polarized. On one hand, some experiments suggests that due to absence of vision, blind and visually impaired people have problems in judging distance using non-visual cues and that vision is necessary for spatial cognition (Eimer, 2004; Pasqualotto & Proulx, 2012). On the other hand, some argue that blind people are as good as sighted in their spatial skills (Schmidt et al, 2013; Tinti et al., 2006) and even better at spatial skills needing non-visual senses, compared to sighted people (Collignon et al. 2009; 2009a; Lederman & Klatzky, 2009). Van der Stoep et al., (2017) also suggest that the absence of vision does not have any influence on the other senses as vision is independent from other senses.

From an early age blind individuals show certain biases and impairments when using tactile and haptic perception to perceive spatial properties of an object, showing larger deviations and errors compared to sighted individuals (Cappagli & Gori., 2016; Cattaneo et al., 2011; Gaunet & Rossettti, 2006; Gori et al., 2021; Landau et al., 1984; Landau, 1991; Postma et al., 2008; Röder et al., 2004). Nevertheless, Kennedy, Gabias and Heller (1992) argue that, although blind individuals show more variable and slower responses when solving spatial tasks, their "abstract framework for space may use the same principals" (pp.175) as sighted individuals (see also Bedny & Saxe, 2012; Cattaneo et al., 2008; Heller& Kennedy., 1990). Furthermore, Papadopoulos and Koustriava (2011) suggested that, although the performance in spatial coding and representations of near space might differ between sighted and blind individuals, there is no difference between blindfolded-sighted and blind participants in their effective use of haptic strategies. On the other hand, Fiehler, Reuschel and Rösler (2009) propose that when using proprioception, congenitally (early) blind people have better spatial discrimination acuity than late blind individuals and the same level of acuity as sighted individuals, arguing that non-visual experience can compensate for loss of vision (also see Lederman & Klatzky, 2009).

Using a tactile detection task, Collignon et al. (2009) also showed that, lateralization of] tactile stimuli by early blind participants is superior to blindfolded-sighted and late blind individuals. They attribute this to the lack of automatic external remapping which would come with visual information, leading to a strong body-centred frame of reference. Previous research on reference frames in blindness commonly suggests that visual deprivation tend to lead to more of an egocentric, body-based spatial representation as opposed to extrinsic, allocentric representations of space which might come with visual input, both in small- and large-scale space (Iachini, Ruggiero & Ruotolo, 2014; Pasqualotto, Lam & Proulx, 2013; Struiksma, Noordzij & Postma, 2011) even when participants are highly familiar with the environment (Corazzini et al., 2010). In a haptic object location task, Coluccia, Mamarella and Cornoldi (2009) indicated that sighted and blind individuals perform similarly when the task requires them to use an egocentric reference frame for space within reach (namely peripersonal space), but the performance of blind individuals deteriorated when they had to use an allocentric reference frame. However, Martolini et al. (2021) recently argued that absence of vision does not alter allocentric spatial coding *per se*, but rather influences blind individuals' (children aged between

6 to 13 years) ability to switch perspective in the haptic domain (also see Crollen & Collignon, 2012).

In absence of vision, blind individuals also rely on their sense of hearing to perceive distance and to navigate. Previous literature on auditory perception of blind and visually impaired people suggests that vision acts as a calibrator for the sense of hearing (and touch) and in the absence of visual information auditory spatial perception is distorted (Aggius-Vella et al., 2019; Gori et al., 2014; Kolarik et al., 2015; Voss, 2016; Wnuczko & Kennedy, 2014). For example, Kolarik et al. (2017) suggested that auditory space is "compressed" in blind people due to a lack of visual information. On the other hand, manipulating the reachability of an object via sound and vision, Ricciardi et al., (2017) found that when detecting an object within reach, congenitally blind people did not show any difference in spatially mapping objects when compared to sighted individuals. Additionally, the same level of performance in detecting objects within peripersonal space was seen even when the object was outside of the blind individual's peripersonal space but within the peripersonal space representation of others. Ricciardi et al (2017) then argue that the peripersonal space of blind individuals develops independently from vision. Furthermore, rather than showing deficits in auditory localisation, due to high reliance on auditory stimuli in spatial perception (e.g., via echolocation), some have argued that blind individuals are better in using sounds to localise objects than sighted individuals (Battal et al., 2020; Röder et al.,2000).

4.2 Communication in blindness

As discussed earlier, communication is a multimodal phenomenon and deictic communication is inherently multimodal. Words we use, our gestures and gaze are expressed in concert as we communicate. Given that people with total visual impairment have no visual access to what we are talking about, either in terms of what a person is talking about or to non-verbal expression, a blind person's understanding will heavily depend on what they hear, such as our description of an object and its position (Feretti, 2008). Therefore, their understanding and communication of space might be more verbally detailed and verbal communication skills would be more specialised than sighted individuals (Galiano & Poerailer, 2011; Sarberg, 2010). Although there is not a large body of research looking into how (early) visual deprivation would influence the choice of spatial demonstrative words, it has been suggested that, from an early age, there is a developmental difference between blind and sighted children in deictic communication, both in terms of gesture and word use (Iossifova et al., 2013; Iverson & Goldin-Meadow, 1997; Iverson et al., 2000; Perez-Pereira, 1997; Röder et al., 2000; Sharkey et al., 2000). For example, Iverson et al. (2000) suggested that during language acquisition, blind children tend to produce less gestures compared to their sighted peers. Similarly, Iossifova et al. (2013) suggested that when pointing towards a specified direction (e.g., in front, behind, up, down, left-right) blind children are slower at pointing than sighted children.

Blind individuals also differ from sighted individuals in language processing. For example, Röder et al., (2000) suggest that blind individuals process language faster than sighted individuals. Pasqualotto, Lam and Proulx (2013) further report that blind participants recall semantically related words better and have an enhanced verbal ability, although early research with blind children indicates difficulties and delays in language acquisition (Anderson, Dunlea & Kekelis, 1984; 2016; Bigelow, 1987). Yet the effects of not having vision tends to reduce as children get older (Rosel et al., 2005). Perez-Pereira (1999) also suggested that blind children (although the sample consisted only of five children) start to use reference terms such as personal pronouns as early as sighted children although the use of demonstratives was infrequent (see also Mills, 1983 and Warren, 1994). Looking at deictic term use and comprehension of a group of blind Persian speakers, Monfared and Estaji (2011) argued that they had difficulties in perception and use of deictic terms. They also suggested that blind individuals need to have a prior interaction with the referents either haptically or auditorily, therefore they tend to use deictic terms less often to refer to novel objects in their surroundings.

Regarding non-verbal communication in visual loss, gesturing during speech tends to be resilient to absence of vision. Blind people gesticulate as they speak or to indicate, even when they know/think that the listener does not have access to this information (Iverson et al., 2000; Iverson & Goldin-Meadow, 1998;2001). This could be due to the articulatory function of gesturing as speakers are not aware of their gestures, but gesturing helps the speaker express their thoughts (Pavelin-Lešić,2010). Blind individuals also use referential gesture such as pointing, although gesture form and frequency might be different from sighted individuals (Gugo & Geld, 2017; Kane, Wobbrock & Ladner, 2011; Sharkey et al., 2000). Recently, Mamus et al., (2023) suggested that blind participants used pointing gestures more often that sighted individuals but used iconic gestures less when describing motion events, although there was no difference between blind and sighted participants for overall use of gestures.

To sum up, we can argue that blind individuals' spatial perception as well as communication show some similarities but also differences to sighted individuals. The findings are not consistent across studies, making it difficult to draw solid conclusions regarding the spatial perception-communication interplay. Such inconsistencies may in part be due to differences in experimental designs as well as between individual factors such as Braille capacity and use of assisting devices during testing (Cattaneo et al., 2011; Schmidt et al., 2013). We could then argue that more standardized methods and research with a wider population of participants with visual impairment is needed to bring a better understanding to (deictic) communication in blindness. Next, I describe what one might regard as a stepping stone on this journey.

4.3 Experiment 6

To bring a better understanding to the non-visual aspects of deictic communication, we utilised the haptic adaptation of the Memory Game paradigm explained earlier with blind and visually impaired individuals. The entire stimuli set up (including the rail and the objects) were portable (Figure 3.1) and was transported to the place of data collection (Great Yarmouth and Norwich hubs of Vision Norfolk). The same experimenter who collected data in Chapter 3 also collected all the data from these participants. After the experiment, participants were asked about their experience with the task. Their responses were recorded and integrated to observations described below. Overall, participants thought that the task was a memory task, similar to the general trend in Memory Game Paradigm experiments, and no-one guessed the aim of the study. However, participants expressed difficulties with sticking only to three words when describing object positions. As we report below, as a group, visually impaired individuals showed a tendency to pick the word “that” more often, regardless of the object position.

4.3.1 Data Collection

The data were collected in Vision Norfolk hubs at Great Yarmouth and Norwich. Participants were recruited with help of Vision Norfolk staff from the members of the organisation who were familiar with the environment (as vision Norfolk carries out activities for the members in these hubs). The conditions were kept as similar to the in-lab conditions as possible (the procedure was the same as in Experiment 5). As the in-lab testing at the University of East Anglia, participants all wore noise blocking headphones with white noise to avoid any auditory cues of distance. Interestingly, blind individuals qualitatively reported that white noise made it very difficult for them to complete the task, which was not so strongly noted by the sighted individuals doing the same task. This would most probably indicate higher reliance of blind individuals to sound compared to sighted, but it is not possible to draw these conclusions since experience with the white noise was not a variable but a mere control for this task.

All the participants had access to an information sheet, instruction sheet and consent form before they attended their scheduled data collection session. They signed consent under the supervision of members of Vision Norfolk and were compensated for their time. They were aware of their right to withdraw should they not wish to continue the experiment and they were made aware that if they chose to stop the experiment, they would still be compensated for their time. Their transport to and from the hub was also compensated.

4.3.2 Participants

In total, we collected data from 10 participants with varying levels of visual loss (If participants had any light vision, they wore blindfolds when they completed the task). We will consider participants both individually and as a group when discussing findings. The sample consisted of 4 female and age range between 50 to 74. Please see Table 4.1 for a full breakdown of participant information.

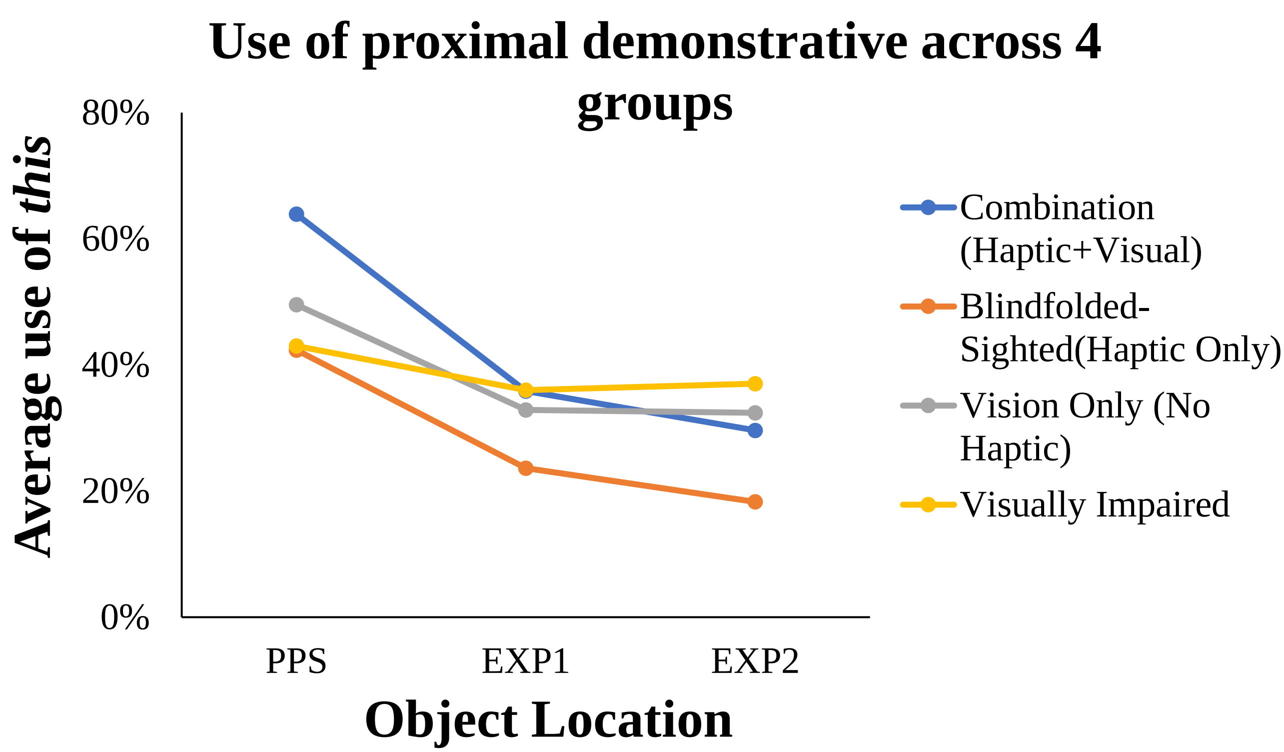
Table 4.1: Demographic information of the blind and visually impaired participants.

ID	Age	Gender	Handedness	Onset	Braille fluency	Employment Status	Education
1	48	Male	Right	Late	Fluent	Volunteering	College
2	50	Female	Right	Born with impairment	Beginner	Volunteering	B-TEC
3	60	Male	Left	Age 16	Fluent	NA	B-TEC
4	67	Male	Right	Congenital	NA	Volunteering	College
5	58	Male	Right	25 (Usher syndrome)	Beginner	Retired, used to work in construction	6th Form
6	74	Female	Right	50	NA	Retired, had a newspaper shop	Age 15
7	31	Male	Ambidextrous	Congenital	Fluent	Volunteering	University
8	61	Female	Ambidextrous	Congenital	Fluent	Retired, worked as programmer	University
9	64	Male	Left	Lost eyesight at 15, difficulties from birth	Fluent	Retired, worked as mechanical engineer	Age 13
10	56	Female	Right	Difficulties from age 35-40	Fluent	Currently unemployed, worked as bank cashier	Until 19

4.3.3 Visually impaired participants as a group

Overall, blind and visually impaired participants picked the word *that* more often, for 68.6% of all responses. Within peripersonal space, blind participants used the word *this* only 43% of the time, in the extrapersonal space 1 36% of the time, and in extrapersonal space 2, 37% of the time (Figure 4.1) Although as a group there is a gradual drop in use of word *this* from peripersonal to extrapersonal space, they used the word *this* in peripersonal space less than half of the time. When we observe the performance of sighted individuals in Chapter 3 (which suggested a significant effect of object position but no between-condition difference), we can get some insight into the differences between different groups while being cautious that visually impaired participants gave differing number of responses and had different levels of sight loss (Table 4.2).

Figure 4.1: Average choice of word *this* by blind/visually impaired participants.



For the sighted-blindfolded group in Experiment 5, the likelihood of choosing *this* was 42% in peripersonal space, 24% in extrapersonal space 1 and 18% at extrapersonal space 2. In the vision only condition, these numbers change to 50%, 33% and 32% respectively. However, the vision + touch condition has the sharpest shifts between regions with likelihood of using this at 64% in peripersonal space, 36% in extrapersonal space 1 and only 30% in extrapersonal space 2. By observing these numbers, we see a similarity amongst sighted and visually impaired groups when compared to the vision-only condition, followed by the blindfolded-sighted condition. Therefore, visually impaired individuals showed the most similarity in demonstrative choices when compared to the condition in which sighted participants only used vision to perceive space. However, to make direct conclusions we need to conduct research, including controlled and matched sample populations of sighted and visually impaired individuals; for now, we can merely report observed similarities and differences in overall behaviour. Moreover, the individual differences between visually impaired participants in performing the task suggested that the method may need to be fine-tuned to be easier for participants without vision to follow. Hence these preliminary findings can be regarded as a stepping stone to future studies, rather than being conclusive in their own right.

4.3.4 Exploring between-participant differences

For the standard experimental procedure, the total number of responses per participant is 48. With the blind and visually impaired sample (N=10), the responses decreased across blocks and only two of the participants were able to complete the full task. Across the 10 participants, the total number of responses was only 287 (out of a possible total of 480 data points for full completion of the task by everyone). Generally, participants were more likely to use the word *that* more often although differences between individuals' preferences for the words were observable Table 4.2 and Figure 4.2.

Table 4.2: Total number of responses collected from each participant.

	Object Position														
	Peripersonal				Extrapersonal 1				Extrapersonal 2					Demonstrative Use	
ID	A	B	C	D	E	F	G	H	I	J	K	L	Total	This	That
1*	2	2	2	2	2	2	2	1	2	2	2	2	23	23(100%)	X
2	4	4	4	4	4	4	4	4	4	4	4	4	48	10(21%)	38
3**	3	3	3	2	3	3	3	2	3	3	3	2	33	16(48%)	17
4	4	4	4	4	4	4	4	4	4	4	4	4	48	4(8.3%)	44
5*	3	2	2	2	3	2	2	2	3	2	2	2	27	X	27
6*	1	1	X	X	1	1	X	X	1	1	X	X	6	X	6
7*	2	2	2	2	2	2	2	2	2	2	2	2	24	2(8.3%)	22
8*	2	2	2	2	2	2	2	2	2	2	2	2	24	21(87.5%)	3
9*	1	1	X	X	1	1	X	X	1	1	X	X	6	6(100%)	X
10	4	4	4	4	4	4	4	4	4	4	4	4	48	8(16.6%)	40
Total	26	25	23	22	26	25	23	21	26	25	23	22	287	90(31.4%)	197

X indicates no demonstrative use. Participant stopped following the protocol.

*Participant did not follow the 3-word-rule anymore so testing stopped

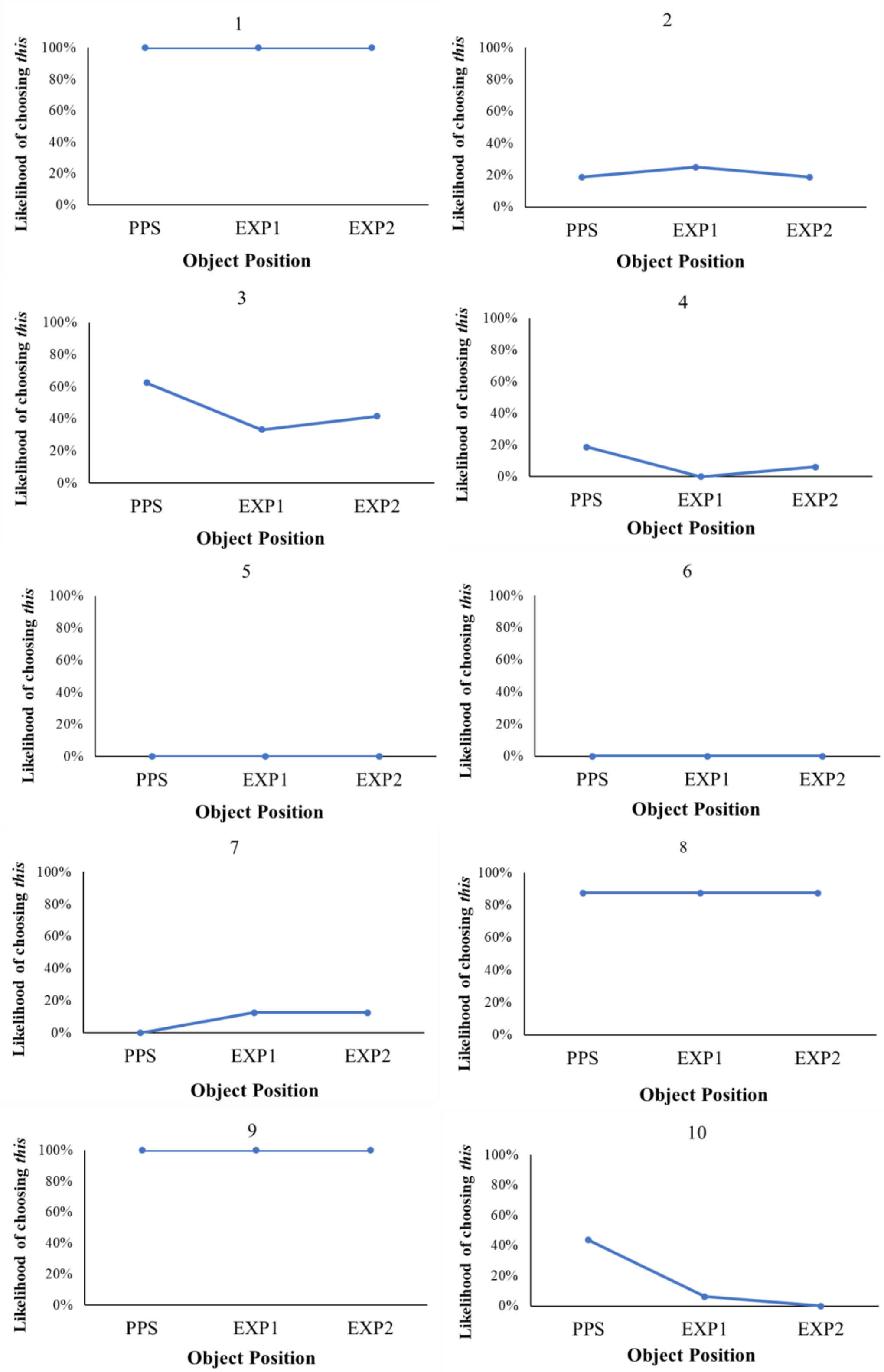
**Participant struggled to complete the task for design reasons

Some of the participants had a preference for one demonstrative and did not use the other throughout the experiment. For example, participants 1 and 9 (Table 4.2), only used the word *this* across all of the object positions while participants 5 and 6 only used the word *that*. Participant 1 switched to the word “the” for rest of the trails after 2 blocks, therefore data collection stopped. This participant further argued that they found demonstrative words to be too “vague” to use in his daily life. When the aim of the study was mentioned to Participant 1 at debrief, he mentioned that the object distance was not an important factor in his word choices.

Participants 7 and 8 were both congenitally blind and only completed half of the experiment and almost had the opposite trend in demonstrative choices. Overall, participant 7 was more likely to use the word *that* on 91.7% of trials completed and participant 86 the word *this* for 87.5% of all trials they completed. Participant 7 mentioned that he found it difficult to stick to 3 words only. He suggested that as he does not use gestures or pointing often to communicate and reported that he needs to use “extra words to explain” himself. Similarly, when asked about word choice strategies, participant 8 mentioned that she “did not see the point of saying *this* or *that*” as she would use words *this* for what happened now and *that* for events of the past with more of a temporal function to demonstratives. When asking about the aim of the task, she mentioned firstly that distance would not influence her word choices and furthermore, since she touched the objects, she would be more inclined to use word *this* regardless of the object’s position. It is likely that space had an influence for this participant’s word choices and as she touched the objects, they felt closer to her deictic centre, and therefore she was likely to use word *this* more often.

Participant 3 was the only participant who used the word *this* almost half of the time on 48% of all trials. He used word *this* more often in the peripersonal space (63% or trials completed), which dropped to 33% in extrapersonal space 1 but increased to 42% in extrapersonal space 2 (but was not as high as in the peripersonal space). This is a trend close to sighted individuals’ performance.

Figure 4.2: Graph representing demonstrative choices of blind and visually impaired participants.



Participants 2, 4 and 10 were the only participants to complete the entire task. All three of these participants were more likely to use word *that* with 79% ,91.7% and 83.4 % use respectively. Participant 2 used the word *this* mostly in the extrapersonal space 1 (25% of the time) and was equally likely to use word *this* in peripersonal and extrapersonal space 2. This participant was mentioned that the word *that* was more readily available for her to use, almost in the word's definite function mentioned earlier. She also mentioned that as she became more familiar and confident with the task, she started to use word *this* when objects were closer and that her demonstrative choice generally depends on if the object belongs to her (ownership effect). On the other hand, participant 4 was more likely to use word *this* with 19% in the peripersonal space did not use word *this* in extrapersonal space 1 and only 6% of the time at the extrapersonal space 2. Participant 10 showed the strongest shift in demonstrative choice. Although they only used word *this* 16.6% of total number of trials, the likelihood of using word *this* was highest in the peripersonal space with 44%, followed by 6% in extened peripersonal space and they did not use word *this* in extrapersonal space. When asked about her word choice strategy, this participant mentioned that she picked word *this* to objects which were closer to her, nevertheless as we can see at the Table 4.2, she did indeed use word *this* more often in the peripersonal space than any other position, but still the likelihood is less than 50% of the trials.

4.4 Concluding remarks

Previous research supported both similarities and differences across senses regarding the perception of space within peripersonal space. For example, vision is argued to calibrate other senses. Based on results reported above, in Chapter 3 with blindfolded-sighted individuals can vaguely guide us into understanding potential implications of lack of visual information to demonstrative use-which showed that there was no difference between groups as a function of sensory input (although there was an observable advantage of vision with haptic condition). Based on these findings, we might argue that individuals with acquired or late onset blindness might use demonstratives with near-far distinction. These findings support the idea that haptic and visual perception show similarities in perception of space, leading to similar choices of demonstratives within and outside peripersonal space. When we introduced a population which has much more experience with absence of vision, the blind and visually impaired individuals, (although sample was small therefore results are not conclusive) in Chapter 4, observed behaviour imply no clear indication of spatial distinction by blind and visually impaired individuals compared to research on sighted individuals. Blind and visually impaired rely either on auditory input, guidance of a sighted individual or remember object location through a memory acquired by previous contact. Although in number of ways the way distance is perceived different from sighted, blind and visually impaired individuals still have knowledge of near and far comparable to the sighted. Therefore, in their daily lives, if blind individuals use demonstratives spatially, we might expect their word use to show near-far distinction of demonstrative choices, similar trend with the research on sighted individuals.

, In Experiment 6, we could observe a similarity in the demonstrative choices between visually impaired group and sighted participants when they relied only on one sense, particularly blindfolded sighted participants. These three groups were more likely to use word *that* overall compared to sighted individuals who used a combination of vision and haptics, but still showed a spatial gradient in their demonstrative choices.

Taking the literature reviewed and results of studies reported throughout this thesis, we could argue that the frequency in the spatial use of demonstratives might be differ between blind/visually impaired individuals and sighted individuals. As reviewed above, the research on perception of distance as well as spatial communication are yet to be conclusive. Perceptually, some research indicate difference between blind and sighted, for example, shrinking of space in blind individuals, while others indicate similar cognition of distance between groups. Linguistically, previous research implied how blind and visually impaired individuals are more descriptive in their use of spatial descriptions compared to sighted, for example, when giving directions to a place or an object (Mamus et al, 2023; Struiksma et al., 2011). Although Experiment 5 did not yield a difference in demonstrative use as a function of restricted vision, we do not have information on the use of demonstratives in a more naturalistic setting, without specific instructions to follow, such as the three-word-rule as in memory game paradigm. Instead, as limited but informative piloting with blind and visually impaired individuals in Experiment 6 yields, blind and visually impaired individuals might choose an elaborate description of object location, with use of, for example, precise coordinates and relative position of the referent in a given context, which might make spatial demonstratives redundant. We could argue this based on individual experiences of the participants in Experiment 6 while attempting to complete the task, as only a few finished the experiment and almost all of the participants reported difficulty in using three words to describe object positions and did not always report the near-far distinction in their demonstrative choices, if they would use these words at all. Further research might even suggest that blind individuals use demonstratives only non-spatially for example with a temporal function as demonstratives are used to address events in the past, present and future, use word *that* in its definite function or only the word *this* for objects they touch to feel their location with no distinction of near and far. The difference in the way demonstratives is used by blind and visually impaired might not only be a perceptual matter or due to lack of space effects on demonstratives, but communicative difference altogether where potential differences process of conceptualisation of space, influence the linguistic expression of spatial relations. Furthermore, absence of vision can have implications to the multimodal nature of spatial demonstratives use. Blind and visually impaired individuals do not have access for the gaze information of the other interlocutors, and they cannot use gaze to guide attention although they might use head movement and orientation comparable to sighted, while they use deictic gesture although they have no access to pointing information unless they are in physical contact with the speaker. Hence, absence of gaze might lead to difficulty in use and understanding of demonstratives spatially.

Overall, the field needs systematic research of communication by populations with differing abilities to bring better theoretical understanding to the field but most importantly to ensure an inclusive society as well as research.

Chapter 5 General Discussion

The motivation for this PhD was to arrive at a better understanding of the fundamentals of deictic communication, focusing on two aspects: multimodality of demonstrative use, and implications of multisensory perception in peripersonal space to the mapping of space to demonstratives. Below I discuss the findings across experiments and their broader implications.

4.5 Multimodality of deictic communication:

The first four studies reported in Chapter 2 shed light on the multimodal nature of communication, bringing a deictic take on the processing of co-speech gesture and gaze. Early research in co-speech gesture usually focused on conventional gestures (such as iconic gestures) where gesticulation is highly associated with the meaning of speech (Kelly et al.,2010; 2015). Here we consider the relative importance of components of deictic communication and elaborate on the possible advantages of seeing multiple modalities over language alone.

When we consider the relative importance of language use and pointing, the results reported overall suggest an advantage of having two modalities as a cue for target location. Previous research with non-deictic gestures suggested that conventional gestures, such as thumbs up gesture, facilitates language processing (Andric & Small, 2012; Cuevas et al., 2019; He et al., 2015). Based on the results reported above, we could argue that gestural deictic reference also has the potential to facilitate processing of a deictic utterance (similar to Bangerter, 2004; Cooperrider, 2016; Stevens & Zhang, 2014). While our experiments only looked at the differing levels of likelihood and decisiveness as a function of available modalities of deictic expressions, in the future, integrating measures of language processing such as response time and predicted accuracy in referent detection (as in Holler et al., 2014 and Kelly et al., 2015). These adaptations could provide insights into whether the presence of multimodal deictic cues facilitates the processing of the deictic utterance. Our findings also indicate that, when pointing and demonstratives mismatch, participants were guided more by the pointing gesture than spatial demonstratives, implying dominance of gesture use over language in directing attention. Potential theoretical reasons for this were discussed above (see Chapter 2), for example, more reliance on gesture from an early age over language (Arslan, Göksun & Nozari, 2023; Grassmann & Tomsello, 2020). Similar to investigating the facilitative effect of gesture, future work should also consider whether incongruity between language and gesture impedes processing and results in longer response times. The experiments reported above consistently indicated pointing to be the dominant cue of reference also when compared to eye gaze, which aligns with the literature suggesting pointing as a more robust and less ambiguous cue for directing attention (Butterworth & Ikatura, 2000; Cooperrider, Slotta & Núñez., 2018). Based on our results, we can argue that participants prioritized pointing over gaze because pointing is an overt communicative signal, indicating a specific target referent (Cooperrider, 2023). In contrast, gaze does not only focus on a referent but also shifts, for example, to unrelated positions in the environment and between interlocutors during face-to-face interactions, potentially making it a less reliable source of information when locating a target. Furthermore, we could argue that (based on previous research and reported findings) eye gaze plays two roles in deictic communication and functions of gaze vary in their primacy: a primary function of modulating social cognition and a secondary function as a cue for referent position. In the first three experiments, eye gaze was only manipulated in its secondary function, where the agent directly looked at the object without sharing gaze. Only in Experiment 4 did we introduce a neutral gaze, during which the agent and the participant shared gaze, although briefly. Consequently, there was an increased effect of eye gaze in Experiment 4. Therefore, when talking about the role of eye gaze in deixis, we should consider its varying functions.

As another factor to consider when interpreting the role of eye gaze in our studies, we should note that we manipulated eye gaze in a very isolated manner, e.g., restricted head movement and no head orientation information, which could have hindered the effect of eye gaze and decreased its salience. Previous research suggests head and gaze orientation and even eyebrow movement together with eye gaze influence the processing of an utterance (Hanna, Brennan & Savietta, 2020; Hömke, Levinson & Holler, 2022; Langton, Watt & Bruce, 2000). By integrating head movements which are highly related to eye gaze, we could disambiguate it as a cue for reference to equate its precision and salience at a comparable level to gesture use.

It should be noted that, in first four experiments utilising online experimentation platform Gorilla, we asked participants to consider which demonstrative/noun the agent would utter based on the image or videos they saw (demonstrative in Experiment 1 or as a noun referring to one of the two potential referents in Experiments 2,3 and 4). Use of such methodology still involves perception and prediction of linguistic reference rather than direct use by the participants. There might be different processes involved in use of demonstratives as a function of predicting what agent would utter versus first-person use by the participants, considering their own perspective and choice. In future, stimuli with first person perspective can be used to immerse participants with the stimuli and induce use of demonstratives from first person perspective. Another drawback of the paradigm described in Chapter 2 is the response options where participants had binary options to of demonstratives or a noun labelling one of the two potential referents.

Compared to real-life situations with large number of word choices, having only two choices could be limiting especially in Experiment 1 where participants could only choose a demonstrative while in our daily lives we might use more (or less) then two words to describe spatial relations, or we use no words altogether. An adaptation of these experiments with an open-ended response can make the design more naturalistic. Further factor undermined in the paradigms reported in Chapters 2 regards the potential effects of social context and interactions on use and understanding of demonstratives. As previously argued, presence (and position) of other interlocutor(s) might influence the use of demonstrative by the speaker. In our experiments, there only was one interlocutor, the agent. The potential effect of other interlocutors on demonstrative could then influence where the attention was drawn to and the shared conversational space was created (see Peeters et al., 2020 and Rubio-Fernandez, 2023). We also did not simulate any prior interaction at the beginning of the experiments to mimic a dyadic context between the agent and the participants although participants were introduced to the task and the agent before data collection, and in Experiment 4 we used stimuli with a shared gaze at the beginning and end of deictic utterance. In other words, future research should factor in the dyadic nature of demonstrative use in social context to increase the ecological validity of the findings.

As we interpret the relative importance of deictic expressions, one factor to consider is the interactions of the participant with the stimuli. There is undoubtedly more to the *process* of choosing a referent than just the choice itself. The way expressions are *observed* can be informative in a number of ways. For example, we do not know *where* participants looked at first; eyes, gesture or linguistic cue, which could inform us about which expression catches attention first, or if people would show attention to any other deictic expression at all when pointing is present. We also do not know *how long* each expression was focused on and *how many times* participants shifted their gaze from one deictic cue to another. The trajectorial information on eye movements could inform us about the order of processing deictic cues while the duration of fixation on each cue could potentially enrich the current findings on the relative importance of each cue.

We also do not know enough about how deictic expressions are *produced* in a dyad. Although there is some previous research such as Bangerter (2006), Cooperrider (2014) and Gonseth et al. (2013) we are in need of more naturalistic experiments observing use of deictic communication in a dyad, such as advanced confederate studies, integrating mixed media of technologies including mobile eye tracking and motion tracking (similar to Holler, Kendrick & Levinson, 2018). Such innovative methodologies can give us a vast amount of information regarding gesture use, gaze following as well as implications of other behaviours in conversation such as turn taking to deictic reference.

4.6 Cross-linguistic variations in deixis

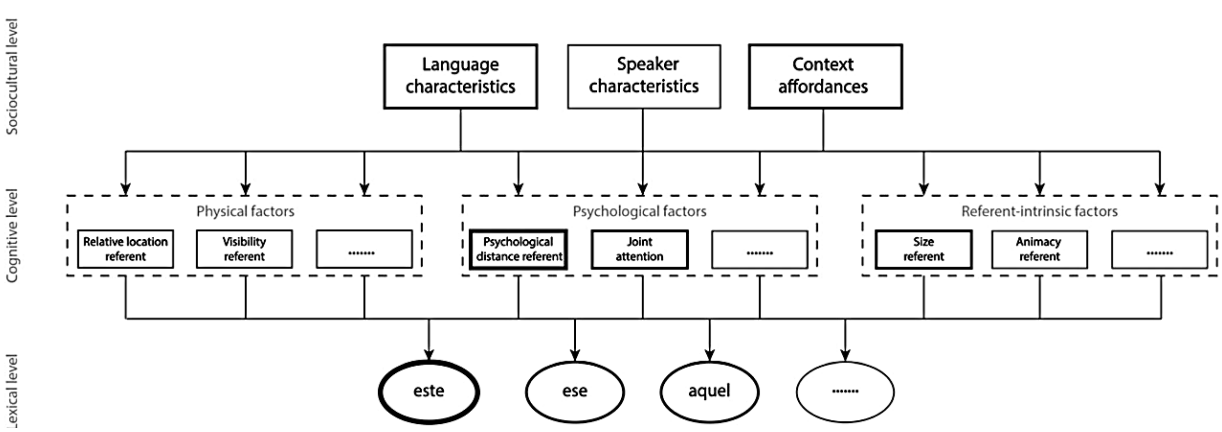
One limitation of the present studies is that the sample population only included monolingual native speakers of English. It is important to be cautious when generalising these findings to languages with different demonstrative systems. For example, in some languages, the use of pointing and eye gaze tends to be obligatory with the demonstrative use. One would wonder what this reliance on gesture and gaze in some languages would imply for the relative importance of deictic expressions. If we run the reported experiments with other languages which have obligatory gesture and gaze, would words have different functions? Would the non-verbal cues have differing relative importance compared to English?

Another cross-linguistic difference across demonstrative systems is the weighting of factors influencing demonstrative use. Although recently Coventry et al. (in press) have shown the mapping between reachability and proximal and distal demonstratives across 29 languages, they also reported differences between languages. Indeed, Peeters et al. (2021) argue that languages might differ in the weight they place on the determinants of demonstratives.

If we look back at their framework for the determinants of demonstratives, Peeters et al. (2021) suggest how social, cognitive and linguistic factors influence demonstrative utterances. They argue a top-down modulation by these factors, starting from the sociocultural factors of speech context, followed by cognitive processes undertaken by the speaker and finally the demonstrative system of a specific language (lexical level). As an example, they apply the framework to the case of the Spanish language, based on the work of Jungbluth (2003). According to Jungbluth (2003), "speakers create a shared conversational space" during their face-to-face conversation and anything within this space is labelled with a proximal demonstrative. Peeters et al. (2020) then pit the findings of Jungbulth (2003) against Coventry et al. (2008), comparing the created shared space of interlocutors versus object nearness as the determinants of demonstrative use. They argue that the memory game paradigm used in Coventry et al. (2008) makes the object's physical distance a "highly salient" (pp.9) factor which influences demonstrative choice while in Jungbluth (2003) the use of demonstratives is more naturalistic where demonstrative use would be framed within a "social and collaborative process" (pp.9. see also Peeters, Hagoort & Özyürek, (2015) and Peeters & Özyürek, (2016)). Based on these two contrasting arguments on use of demonstratives, Peeters et al. (2021)

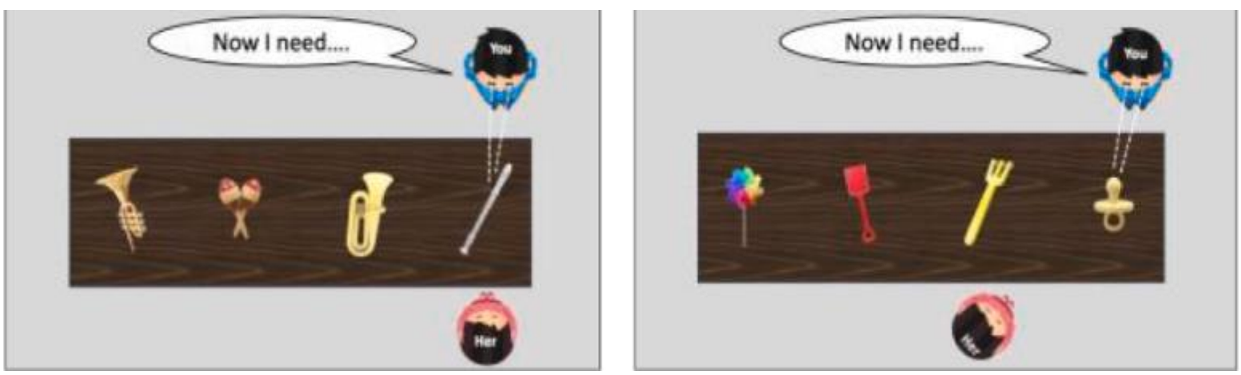
conclude that the effect of psychological and physiological factors influencing the word use in Spanish can be contextual, varying between experimental set-ups manipulating referent position and naturalistic dyads with face-to-face conversation. When applying these findings to their framework, in a dyadic, face-to-face context as in Jungbluth (2003), the relation of factors leading to the deictic utterance would be shown as in Figure 5.1. The speaker would adapt their demonstrative use based on the psychological availability of the listener and the shared space they create where psychological distance would outweigh the physical proximity ("physical factors" at "cognitive level" in Figure 5.1) when choosing a demonstrative. Therefore, Spanish speakers' use of demonstratives would be more influenced by psychological factors, such as space created between interlocutors, than physical factors such as referent location.

Figure 0.1: Application of the framework of Peeters et al. (2021) to use of Spanish demonstratives.



Developing on the contrasting views around the role of psychological versus physical factors influencing demonstrative use in different demonstrative systems, Rubio-Fernandez (2022) investigated the determinants of demonstratives in four languages: one with a two-word demonstrative system (English) and three others with three-word systems (Spanish, Catalan and Japanese). In this study, the author modified the classical memory-game paradigm to an online environment (Figure 5.2), varying the relative positions of the interlocutors as well as the position of the target referent. The author categorises languages as distance-oriented and person-oriented systems. In distance-oriented demonstrative systems, the proximity of the referent influences the demonstrative choice, which has been previously argued to be the case especially for two-word demonstrative systems. In person-oriented systems, the speaker's demonstrative choices are shaped by the position of the listener, which has been suggested more often in three-word systems (but also in two-word systems such as English and Dutch).

Figure 0.2: Example of the stimuli used in Rubio-Fernandez, 2022.
In this experiment, the position of the interlocutors varied. Participants were asked to fill in the sentence in the speech bubble with a demonstrative. Number of options would vary respectively to the language the participant spoke.



The results from Rubio-Fernandez (2022) suggested a number of insightful findings. The trend of word choices of Spanish speakers supported the person-oriented view of demonstrative choices. While the proximal demonstrative (*este*) was preferred when the referent was close to the speaker and the distal demonstrative (*aquel*) was preferred for a position farther both from the speaker and closer to the listener, the medial demonstrative form (*ese*) was used more often when the referent was near the listener but farther from the speaker. Therefore, unlike English speakers, Spanish speakers showed sensitivity to listener position. These results contradict the findings of Jungbluth (2003) and are more in line with the results reported by Coventry et al. (2008). Similarly, Japanese participants picked the medial term (*sore*) when the referent was far from the speaker but near the listener and the distal term when the referent was far from both interlocutors (*are*). On the other hand, Turkish speakers did not show sensitivity to the speaker position when choosing any of the demonstratives. In Turkish, two demonstratives, *bu* (proximal term) and *o* (distal term) were used based on the position of the referent, putting the Turkish into the distance-oriented language systems. However, also rather interestingly, the third demonstrative *su* was used for attention correction, with no relation of this word choice to the position of speaker or listener (also supporting previous work of Küntay and Özyürek, 2006 and Özyürek,1998). Therefore, Rubio-Fernandez argues that not all demonstrative systems with three words are person-oriented. These results can further support the framework suggested by Peeters et al. (2021) and expand their claims that determinants of demonstrative are weighted variably in different demonstrative systems.

To sum up, the experiments reported in Chapter 2 need to be further developed in two respects. Firstly, we need a more naturalistic and multidisciplinary approach to the research on deixis to improve the ecological validity and generalizability of the findings. Secondly, the experiments on the relative importance of deictic expressions should be conducted with speakers of other languages with different demonstrative systems to assess if deictic cues are prioritized universally or differ in their weightings across languages. Cross-linguistic replication would answer a number of questions, such as whether speakers of person-oriented languages rely more on gaze, a social cue of joint attention, than speakers of distance-oriented languages (as described in Rubio-Fernandez), and whether speakers of languages where gaze or gesture (supposedly) obligatorily accompany use of specific spoken demonstratives forms show stronger effects of non-linguistic cues as pointers to reference than speakers of languages where non-linguistic cues are more optional.

4.7 Manipulating spatial perception

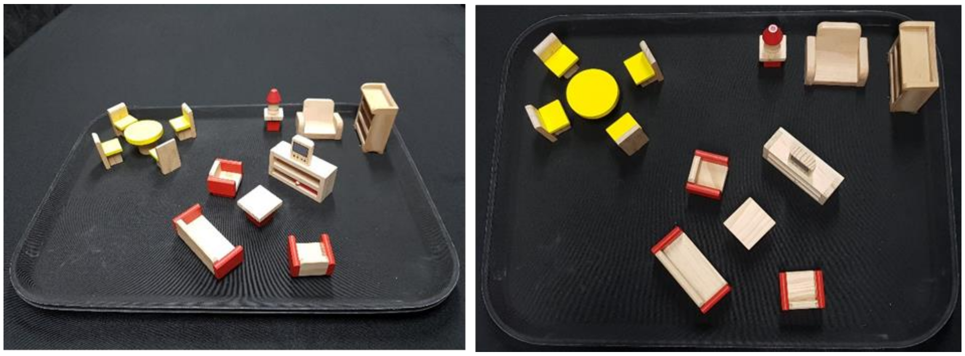
Findings from the second set of the experiments reported above bring a new perspective to the mapping of space to demonstrative choices vis-a-vis consideration of the multisensory nature of peripersonal space. Results from a haptic adaptation of the memory game paradigm reported in Chapter 3 suggest that, regardless of the sense used, there is a distinction between proximal-distal demonstrative use as a function of object position.

Methodologically, it is worth noting that there is room for further optimisation in the design of the experimental protocol to tackle factors which might have hindered any effect of the sensory modality used. For example, the object names and their labels (e.g., textures) might have been difficult to remember for participants (we did not measure object labelling accuracy), and particularly so for sighted individuals. Such cognitive load may have masked possible differences between senses. Additionally, it is possible that some of the object positions were too far to reach to touch, where participants had to show effort (although not too high) to reach the objects in extrapersonal space. The need to move in order to touch the object might have strengthened the effect of object position, eliminating any difference in behaviour as a function of senses used. Reducing object locations in the farther positions could help to refine the effect of sensory perception. Another factor to consider is the cognitive spatial abilities of individuals as well as their physical capacities. By calculating participant-specific positions of objects, we wanted to make each position reachable for participants, based on their physical capacities. However, physical capacities might have not been an effective measure of between-participant variance in spatial cognition, because as discussed below, variance in (cognitive) spatial abilities (Markostamou and Coventry, 2022) tends to influence spatial language.

As briefly touched upon in Chapter 4, most of the work in spatial perception and deixis interplay is based on individuals who have access to visual information. We do not have a comprehensive understanding of spatial communication in blindness; a group of people who rely on non-visual senses to perceive space. There is very limited work in the blindness and visual impairment field on deixis, with only a handful of papers on spatial language use (such as Struiksma, Noordzij and Postma, 2011), and a few targeting multimodal communication directly in naturalistic discourse (such as Iverson & Goldin-Meadow, 2001) or gesture use (Mamus et al 2023). There is a wide range of information we can obtain from individuals with visual loss, including characteristics of gesture and language use in blindness expanding our current knowledge of communication of people with different abilities, all of which would lead to scientific, societal and technological advancements. With the pilot study reported in Chapter 4, we attempted to create an experimental paradigm to investigate demonstrative use in the absence of vision. However, this work needs to be optimized based on insights gained from piloting with blind individuals and improved into a validated and replicated measure of demonstratives.

A project which did not make it to this thesis due to time constraints concerned a haptic adaptation of Tenbrink, Andonova and Coventry (2011) where blind and sighted individuals haptically and/or visually explored an arrangement of dollhouse furniture (Figure 5.3). With use of video cameras and wearable eye trackers, we intended to explore spatial exploration and learning strategies as well as language and gesture use by comparing blind and sighted participants as they recollect object positions. We were also going to measure object location memory. This study would provide valuable insights into the cross-sensory differences in the formation of spatial representation and for the use of language and gesture when describing spatial alignments, and use of such a paradigm should be pursued to help illuminate potential differences in spatial communication with or without vision.

Figure 0.3: The arrangement of stimuli for visual and/or haptic exploration.

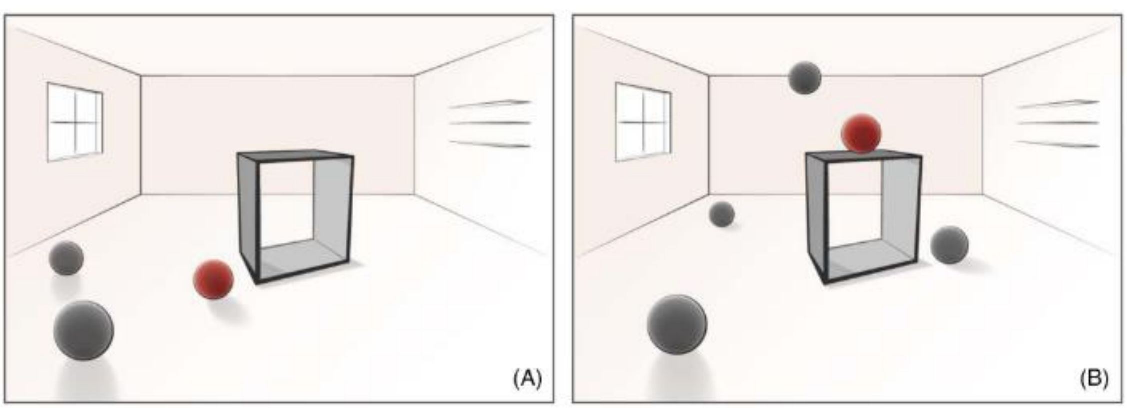


In future research, the study of deixis in the absence of vision would benefit from the use of advanced techniques such as motion tracking and/or auditory VR (such as Project Triton) to better understand the spatial language use and non-verbal co-speech behaviour (not only gesture but head and body orientation). It should also be noted that most of the work on spatial perception in the absence of vision has been based on auditory perception. In the studies reported in Chapters 3 and 4, we used haptic spatial perception as a contrast to visual perception. In the future, the use of auditory perception could offer a reliable, informative and complementary means with which to investigate the relation between non-visual spatial perception and deictic expressions.

4.8 Spatial abilities and language

As we already established, spatial perception plays a fundamental role in the use of English demonstratives, with a mapping between reachable/non-reachable space and the use of *this* and *that* (e.g., Coventry et al., 2008, 2014). However, it is also the case that individual differences exist in how demonstratives are used within languages, including English (Coventry et al., in press). Recently, Markostamou and Coventry (2022) carried out a novel assessment of spatial language production (the Spatial Naming Test) across the lifespan, together with standard visuospatial tasks (such as mental rotation, visuospatial reasoning and visual organization) and verbal tasks of a non-spatial nature (such as action naming and object naming) to investigate the relationship between of visuospatial abilities, verbal abilities and spatial language production across the lifespan. During the Spatial Naming Test, participants were asked to use prepositions to describe the position of "the red ball" (a geometrical object) in relation to "the cube" (see Figure 5.4).

Figure 0.4: Examples of the stimuli used in Markostamou & Coventry, 2022.



Markostamou and Coventry (2022) found a general decline in spatial naming in older adults. The results suggested that the use of prepositions was predicted by age *and* visuospatial abilities (but not by verbal ability). Interestingly, overall visuospatial ability predicted performance on the Spatial Naming Task regardless of age where better visuospatial abilities predicted more accurate spatial naming performance on the spatial naming task.

The findings of Markostamou and Coventry have two main methodological implications for the experiments described throughout this thesis. Firstly, the experiments reported in Chapters 2 and 3 mostly consisted of a younger population of university students, therefore, reporting findings in these chapters might not be generalisable to a wider population, if ageing does indeed mediate the relation between visuospatial abilities and spatial language. In the future, we could consider widening the sample and applying measures of spatial ability to our studies. Second, Markostamou and Coventry (2022) indicated visuospatial ability as a significant factor in the spatial naming task regardless of age. By considering spatial ability as a concept in studies concerning deixis, we could test whether variation in spatial abilities also affects peripersonal space representation, and therefore demonstrative choices, or alternatively whether the mapping of demonstratives to peripersonal space is robust and not subject to individual differences in spatial abilities. If spatial abilities would indeed contribute to the interplay of peripersonal space and demonstrative use, it might be a stronger factor than the sensory domain used and might explain the findings of Experiment 5. Individual differences in spatial ability could also have implications for the research with blind individuals. The studies investigating spatial perception of blind and visually impaired individuals do not usually assess the spatial abilities of the participants, but merely assume their spatial abilities through their years with visual loss and by measuring skills such as navigation, perception of orientation and sound localization. A between-participant difference in spatial abilities could also explain differences in the findings of studies on perception and language in blindness. For this, the field is in need of a reliable and validated neuropsychological assessment of spatial abilities of blind individuals (such as Goodridge et al., 2021; Lopez et al., 2020).

4.9 Neural basis of deixis

Another aspect of deictic communication that has received insufficient attention is its neural underpinnings. As mentioned in Chapter 1, Rocca et al. (2020) is one of the few studies which tried to identify brain regions involved in use of spatial demonstrative (also see Peeters et al., 2015). In this study Rocca et al. had participants listen to a narrative which involved demonstratives, and they observed the hemodynamic response in the parts of the narrative with demonstratives. However, this study has limitations in that it did not directly and exclusively manipulate demonstratives. The responses obtained could have been influenced by the narrative content (although the authors used a number of control analysis in an attempt to counteract any confounding influences). Listening to a narrative alone might not be the strongest measure of demonstratives, especially considering the multimodal nature of demonstrative use. Prior research, while not primarily centred on demonstratives, also indicated enhanced processing of utterances accompanied by gestures (Cuevas et al., 2021). Therefore, considering the application of the above reported Experiment 4 methodology to neuroimaging studies could give us enriched information about the neural correlates of deictic communication.

Although previous work has attempted to illuminate the neural underpinnings of demonstrative use, there are still a number of questions regarding brain networks involved in deixis. For example, if demonstratives are indeed lexicalised with gestures, would we expect activation in (pre)motor areas when processing demonstratives (especially for the speakers of languages where pointing is obligatory)? Would the *social parts* of the brain (see van Overwalle, 2009 for a review) also be involved in deictic expressions, with more likelihood of involvement of these areas in some languages (e.g., person-centred languages) more than others, mirroring the approach in Rubio-Fernandez (2022)? The future development of neuroimaging-friendly methodologies to investigate deixis would be informative regarding the multimodal nature of demonstrative use. For example, by integrating the experimentation of Rubio-Fernandez (2022) (which adapted the memory game paradigm by both digitizing the experiment and systematically manipulating a range of determinants of demonstratives), we can also investigate neural correlates of the determinants of demonstratives by observing any related activation in spatial, social or linguistic parts of the brain.

In line with neural underpinnings, another aspect in which we lack understanding is use of deixis by people with neurodiversity. Using the methodology of Markostamou and Coventry (2020), Bochynska et al. (2020) found that individuals with autistic spectrum disorder tend to have difficulties in use of spatial language. Previous work with children with autism also suggested these individuals have impairments in social communication skills including sharing and following eye gaze (Krstovska-Guerrero & Jones, 2015) and perspective taking (Pearson, Ropar & Hamilton, 2013) which relate to (spatial) language (Beveridge & Pickering, 2013; Tosi, Pickering & Branigan, 2020). One question which is yet to be answered is whether there are implications of autism to deictic expression production and comprehension.

Moreover, we are yet to know whether neurological conditions affecting the perception in peripersonal space would also influence deictic expressions. As I discussed in Chapter 1, individuals with spatial neglect and extinction experience difficulties in the processing of sensory information in peripersonal space. A natural question which stems from this symptom is to look at the implications of these disorders on the use of deictic expressions.

4.10 Conclusion

This thesis has investigated the multimodal nature of deictic communication and the implications of multisensory perception on the use of demonstratives. In large part, the results confirm increases in the decisiveness of referent choice when seeing multiple deictic expressions over exposure to language only, where gestural indication of reference is the strongest cue to reference. Contrasting use of haptic and/or visual senses in perception of space, we found no difference in the mapping of demonstratives to peripersonal space as a function of sense(s) used, further supporting the multisensory nature of peripersonal space and suggesting that the effect of referent position is resilient to the sensory domain used for spatial perception. While these findings are insightful, we need to conduct more multidisciplinary research, looking into spatial, social, perceptual and neuroscientific features of deictic expressions, in order to have a comprehensive understanding of their use.

References

Abashidze, D., & Knoeferle, P. (2021). Influence of Actor's Congruent and Incongruent Gaze on Language Processing. *Frontiers in Psychology*, *12*, 4057.

Aggius-Vella, E., Campus, C., Kolarik, A. J., & Gori, M. (2019). The Role of Visual Experience in Auditory Space Perception around the Legs. *Scientific Reports*, *9*(1), 1–10.

Ahn, D. (2017). Definite and demonstrative descriptions: A micro-typology. *Proceedings Of GLOWin Asia XI*, *1*, 33–48.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and cognitive processes*, *15*(6), 593-613.

Andersen, E. S., Dunlea, A., & Kekelis, L. (2016). The impact of input: language acquisition in the visually impaired: *Http://Dx.Doi.Org/10.1177/014272379301303703*, *13*(37), 23–49. https://doi.org/10.1177/014272379301303703

Andersen, E. S., Dunlea, A., & Kekelis, L. S. (1984). Blind Children's Language Resolving Some Differences. *Journal of Child Language*, *11*(3), 645–664. https://doi.org/10.1017/S0305000900006000

Anderson, S. R., and Keenan, E. L. (1985). *Language Typology and Syntactic Description*, Ed. T. Shopen (Cambridge: Cambridge University Press), 259–308.

Andric, M., & Small, S. L. (2012). Gesture's neural language. *Frontiers in psychology*, *3*, 99.

Arbib, M. A. (2008). From grasp to language: Embodied concepts and the challenge of abstraction. *Journal of Physiology-Paris*, *102*(1–3), 4–20. https://doi.org/10.1016/J.JPHYSPARIS.2008.03.001

Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis* (Vol. 16). Oxford University Press.

Arikan, G. (2018). The language of Space without sight: Trends in spatial demonstrative use and object location memory with haptic perception. Unpublished *Masters Internship Report written to fulfil Requirement of Research Masters Neuroscience and Cognition, Utrecht University, Netherlands*.

Arslan, B., Ng, F., Göksun, T., & Nozari, N. (2023). Trust my gesture or my word: How do listeners choose the information channel during communication? *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0001253

Augestad, L. B. (2017). Mental health among children and young adults with visual impairments: A systematic review. *Journal of Visual Impairment and Blindness*, *111*(5), 411–425. https://doi.org/10.1177/0145482x1711100503

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415–419.

Bangerter, A., & Louwerse, M. M. (2005). Focusing Attention with Deictic Gestures and Linguistic Expressions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *27* (27).

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Berti, A., & Frassinetti, F. (2000). When far becomes near: Remapping of space by tool use. Journal of Cognitive Neuroscience, 12(3), 415–420.

Barutchu, A., Crewther, D. P., & Crewther, S. G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Developmental Science*, *12*(3), 464–473. https://doi.org/10.1111/j.1467-7687.2008.00782.x

Bates, E. (1979). Cognition and communication from 9 to 13 months: Correlational findings. *The emergence of symbols: Cognition and communication in infancy*.

Battal, C., Occelli, V., Bertonati, G., Falagiarda, F., & Collignon, O. (2020). General enhancement of spatial hearing in congenitally blind people. *Psychological Sciences, 31*(9), 1129–1139. https://doi.org/10.1177/ 0956797620935584

Bedny, M., & Saxe, R. (2012). Insights into the origins of knowledge from the cognitive neuroscience of blindness. *Cognitive Neuropsychology*, *29*(1–2), 56–84. https://doi.org/10.1080/02643294.2012.713342

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, *44*(2), 178–190. https://doi.org/10.1016/j.neuropsychologia.2005.05.007

Bertenthal, B. I., Boyer, T. W., & Harding, S. (2014). When do infants begin to follow a point?. *Developmental Psychology*, *50*(8), 2036.

Berti, A., & Rizzolatti, G. (2002). 3.3 Coding near and far space. *The cognitive and neural bases of spatial neglect*, 119.

Beveridge, M. E. L., & Pickering, M. J. (2013). Perspective taking in language: Integrating the spatial and action domains. In *Frontiers in Human Neuroscience* (Vol. 7, Issue SEP, p. 577). Frontiers Media S. A. https://doi.org/10.3389/fnhum.2013.00577

Bigelow, A. (1987). Early words of blind children. *Journal of Child Language*, *14*(1), 47–56. https://doi.org/10.1017/S0305000900012721

Bochynska, A., Coventry, K. R., Vulchanov, V., & Vulchanova, M. (2020). Tell me where it is: Selective difficulties in spatial language on the autism spectrum. *Autism*, *24*(7), 1740-1757. https://doi.org/10.1177/1362361320921040

Bohnemeyer, J. (2018). Yucatec Demonstratives in Interaction: Spontaneous versus Elicited Data. In S. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in Cross-Linguistic Perspective* (Language Culture and Cognition, pp. 176-205). Cambridge: Cambridge University Press. doi:10.1017/9781108333818.009

Bonino, D., Ricciardi, E., Sani, L., Gentili, C., Vanello, N., Guazzelli, M., Vecchi, T., & Pietrini, P. (2008). Tactile spatial working memory activates the dorsal extrastriate cortical pathway in congenitally blind individuals. *Archives Italiennes de Biologie*, *146*(3–4), 133–146. https://doi.org/10.4449/aib.v146i3.780

Bottini, R., Mattioni, S., & Collignon, O. (2016). Early blindness alters the spatial organization of verbal working memory. *Cortex*, *83*, 271–279. https://doi.org/10.1016/j.cortex.2016.08.007

Brozzoli, C., Makin, T. R., Cardinali, L., Holmes, N. P., & Farnè, A. (2011). Peripersonal space: A multisensory interface for body-object interactions. In *The Neural Bases of Multisensory Processes*.

Brozzoli, C., Pavani, F., Urquizar, C., Cardinali, L., & Farnè, A. (2009). Grasping actions remap peripersonal space. *NeuroReport*, *20*(10), 913–917. https://doi.org/10.1097/WNR.0B013E32832C0B9B

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), 1–38. https://doi.org/10.5334/JOC.72/METRICS/

Bufacchi, R. J., & Iannetti, G. D. (2018). An action field theory of peripersonal space. *Trends in cognitive sciences*, *22*(12), 1076-1090.

Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.

Burenhult, N. (2008). Spatial coordinate systems in demonstrative meaning. *Linguistic Typology*, *12*(1), 99–142. https://doi.org/10.1515/LITY.2008.033

Butterworth, B., & Hadar, U. (1989). Gesture, Speech, and Computational Stages: A Reply to McNeill. *Psychological Review*, *96*(1), 168–174. https://doi.org/10.1037/0033-295X.96.1.168

Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology*, *18*(1), 25-50.

Caldano, M., and Coventry, K. R. (2019). Spatial demonstratives and perceptual space: to reach or not to reach. *Cognition 191*, 103989. https://doi.org/10.1016/j.cognition.2019.06.001

Candlin, F. (2003). Blindness, art and exclusion in museums and galleries. *International Journal of Art & Design Education*, *22*(1), 100-110.

Candlin, F. (2004). Don't touch! Hands off! Art, blindness and the conservation of expertise. *Body & Society*, *10*(1), 71-90.

Capirci, O., & Volterra, V. (2008). Gesture and speech: The emergence and development of a strong and changing partnership. *Gesture*, *8*(1), 22-44.

Cappagli, G., & Gori, M. (2016). Auditory spatial localization: Developmental delay in children with visual impairments. *Research in Developmental Disabilities*, *53–54*, 391–398. https://doi.org/10.1016/j.ridd.2016.02.019

Cardin, V., Rudner, M., De Oliveira, R. F., Andin, J., Su, M. T., Beese, L., Woll, B., & Rönnberg, J. (2018). The Organization of Working Memory Networks is Shaped by Early Sensory Experience. *Cerebral Cortex*, *28*(10), 3540–3554. https://doi.org/10.1093/cercor/bhx222

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i.

Caruana, N., Inkley, C., Nalepka, P., Kaplan, D. M., & Richardson, M. J. (2021). Gaze facilitates responsivity during hand coordinated joint attention. *Scientific Reports 2021 11:1*, *11*(1), 1–11.

Cattaneo, Z., & Vecchi, T. (2011). 2 Blindness and Sensory Compensation. *Blind Vision: The Neuroscience of Visual Impairment*, *2004*, 11–48.

Cattaneo, Z., Fantino, M., Tinti, C., Pascual-Leone, A., Silvanto, J., & Vecchi, T. (2011). Spatial Biases in Peripersonal Space in Sighted and Blind Individuals Revealed by a Haptic Line Bisection Paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1110–1121.

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., & Pietrini, P. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience and Biobehavioral Reviews*, *32*(8), 1346–1360. https://doi.org/10.1016/j.neubiorev.2008.05.002

Chandralal, D. (2010). *Sinhala*. Amsterdam: John Benjamins

Chebat, D. R., Harrar, V., Kupers, R., Maidenbaum, S., Amedi, A., & Ptito, M. (2018). Sensory Substitution and the Neural Correlates of Navigation in Blindness. *Mobility of Visually Impaired People: Fundamentals and ICT Assistive Technologies*, 167–200. https://doi.org/10.1007/978-3-319-54446-5_6

Chen, D., & Groves, C. C. (2023). Providing Emotional Support to Parents and Caregivers of Infants and Toddlers With Visual Impairments. *Journal of Visual Impairment and Blindness*, *117*(2), 175–182. https://doi.org/10.1177/0145482X231169721

Chieffi, S., Secchi, C., & Gentilucci, M. (2009). Deictic word and gesture production: their interaction. *Behavioural brain research*, *203*(2), 200-206.

Church, R. B., Kelly, S., & Holcombe, D. (2014). Temporal synchrony between speech, action and gesture during language production. *Language, Cognition and Neuroscience*, *29*(3), 345-354.

Cienki, A., & Mittelberg, I. (2013). Creativity in the forms and functions of spontaneous gestures with speech. *Creativity and the Agile Mind*, 231-252.

Clark, E. V. (1978). From gesture to word: On the natural history of deixis in language acquisition. In J. S. Bruner & A. Garton (Eds.), *Human growth and development*, (pp.85–120). Oxford: Oxford University Press

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language*, *5*(3), 457–475. https://doi.org/10.1017/S0305000900002099

Clark, H. H. (2003). Pointing and placing. In *Pointing* (pp. 251-276). Psychology Press.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123 (3)*, 324–347.

Cleary-Kemp, J. (2007). Universal uses of demonstratives: Evidence from four Malayo-Polynesian languages. *Oceanic Linguistics*, *46*(2), 325–347. https://doi.org/10.1353/ol.2008.0008

Cléry, J., Guipponi, O., Odouard, S., Pinède, S., Wardak, C., & Hamed, S. B. (2017). The prediction of impact of a looming stimulus onto the body is subserved by multisensory integration mechanisms. *Journal of Neuroscience*, *37*(44), 10656-10670.

Cléry, J., Guipponi, O., Wardak, C., & Ben Hamed, S. (2015). Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: Knowns and unknowns. *Neuropsychologia*, *70*, 313–326. https://doi.org/10.1016/j.neuropsychologia.2014.10.022

Collignon, O., Charbonneau, G., Lassonde, M., & Lepore, F. (2009). Early visual deprivation alters multisensory processing in peripersonal space. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2009.07.025

Collignon, O., Voss, P., Lassonde, M., & Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Experimental brain research*, *192*, 343-358.

Coluccia, E., Mammarella, I. C., & Cornoldi, C. (2009). Centred egocentric, decentred egocentric, and allocentric spatial representations in the peripersonal space of congenital total blindness. *Perception*, *38*(5), 679–693. https://doi.org/10.1068/p5942

Cooney, S. M., Brady, N., & McKinney, A. (2018). Pointing perception is precise. *Cognition*, *177*, 226-233.

Cooperrider, K. (2016). The Co-Organization of Demonstratives and Pointing Gestures. *Discourse Processes*, *53*(8), 632–656. https://doi.org/10.1080/0163853X.2015.1094280

Cooperrider, K. (2023, accepted). Fifteen ways of looking at a pointing gesture. https://doi.org/10.31234/osf.io/2vxft

Cooperrider, K., Fenlon, J., Keane, J., Brentari, D., & Goldin-Meadow, S. (2021). How pointing is integrated into language: Evidence from speakers and signers. *Frontiers In Communication*, 6(567774).

Cooperrider, K., Slotta, J., & Núñez, K. (2018). The preference for pointing with the hand is not universal. *Cognitive Science, 42 (4),* 1375–1390

Corballis M. C. (2010). Mirror neurons and the evolution of language. *Brain and language*, *112*(1), 25–35.

Coventry, K. R. Gudde, H. B., Diessel, H., Collier, J., Guijarro-Fuentes, P., Vulchanova, M., Vulchanov, V., Todisco, E., Reile, M., Breunesse, M., Plado, H., Bohnemeyer, J., Bsili, R., Caldano, M., Dekova, R., Donelson, K., Forker, D., Park, Y., Pathak, L. S., Peeters, D., Pizzuto, G., Serhan, B., Apse, L., Hesse, F., Hoang, L., Hoang, P., Igari, Y., Kapiley, K., Khutsishvili, T., Kolding, S., Priiki, K., Mačiukaitytė, I., Mohite, V., Nahkola, T., Tsoi, S. Y., Williams, S., Yasuda, S., Cangelosi, A., Duñabeitia, J. A., Mishra, R. K., Rocca, R., Šķilters, J., Wallentin, M., Žilinskaitė-Šinkūnienė, E, Incel, O. D. (under review). Spatial communication systems reflect universal actual constraints. *Manuscript under review.*

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive Psychology*, *69*, 46–70. https://doi.org/10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near-far perceptual space and spatial demonstratives. *Cognition, 108*(3), 889–895.

de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The Interplay Between Gesture and Speech in the Production of Referring Expressions: Investigating the Tradeoff Hypothesis. *Topics in Cognitive Science*, *4*(2), 232–248. https://doi.org/10.1111/J.1756-8765.2012.01183.X

Cuevas, P., He, Y., Billino, J., Kozasa, E., & Straube, B. (2021). Age-related effects on the neural processing of semantic complexity in a continuous narrative: Modulation by gestures already present in young to middle-aged adults. *Neuropsychologia*, *151*(December 2020), 107725. https://doi.org/10.1016/j.neuropsychologia.2020.107725

Cuevas, P., Steines, M., He, Y., Nagels, A., Culham, J., & Straube, B. (2019). The facilitative effect of gestures on the neural processing of semantic complexity in a continuous narrative. *NeuroImage*, *195*(March), 38–47. https://doi.org/10.1016/j.neuroimage.2019.03.054

Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology Paris*, *98*(1-3 SPEC. ISS.), 249–258. https://doi.org/10.1016/j.jphysparis.2004.03.011

Desrochers, S., Morissette, P., & Ricard, M. (1995). Two perspectives on pointing in infancy. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 85–101). Hillsdale, NJ: Lawrence Erlbaum Associates.

Deutscher, G. (2005). *The unfolding of language: An evolutionary tour of mankind's greatest invention*. Macmillan.

Di Pellegrino, G., & Làdavas E. (2015). Peripersonal space in the brain. *Neuropsychologia*, *66*, 126-133.

Di Pellegrino, G., Làdavas, E., & Farne', A. (1997). Seeing where your hands are. *Nature*, *388*(6644), 730. https://doi.org/10.1038/41921

Diessel, H. (1999). *Demonstratives: Form, Function, and Grammatization*. Amsterdam: John Benjamins Publishing Company.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, *17*, 463–489. https://doi.org/10.1515/COG.2006.015

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space. *Linguistics and Language Compass*, *8*(3), 116–132. https://doi.org/10.1111/lnc3.12066

Diessel, H., & Coventry, K. R. (2020). Demonstratives in Spatial Language and Social Interaction: An Interdisciplinary Review. *Frontiers in Psychology*, *11*, 3158.

Diessel, H., & Monkahov, S. (2022). Acquisition of demonstratives in cross-linguistic perspective. *Journal of Child Language*, *11*, 1-32.

Dixon, R. M. (2003). Demonstratives: A cross-linguistic typology. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, *27*(1), 61-112.

Drijvers, L., & Holler, J. (2023). The multimodal facilitation effect in human communication. *Psychonomic Bulletin & Review*, *30*(2), 792-801.

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. In *Current Biology* (Vol. 10, Issue 20). Current Biology Ltd. https://doi.org/10.1016/S0960-9822(00)00740-5

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. Visual Cognition, 6, 509–540

Edwards, S. G., Stephenson, L. J., Dalmaso, M., & Bayliss, A. P. (2016). Social orienting in gaze leading: A mechanism for shared attention. *Proceedings of the Royal Society B.*, *282*, 20151141.

Eibl-Eibesfeldt, I. (1989). *Human ethology.* Aldine de Gruyter.

Eimer, M. (2004). Multisensory Integration: How Visual Experience Shapes Spatial Perception. *Current Biology*, *14*(3), R115–R117. https://doi.org/10.1016/j.cub.2004.01.018

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behaviour: Categories, Origins, Usage and Coding. *Semiotica*, *1*, 49–98.

Enfield, N. J. (2003). *Demonstratives in Space and Interaction: Data from Lao Speakers and Implications for Semantic Analysis*. *79*(1), 82–117. https://doi.org/10.1353/lan.2003.0075

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferretti, F. (2008). Blindness, visual content and neuroscience. *Accessed March*, *7*, 118-135.

Feyereisen, P. (1987). Gestures and Speech, Interactions and Separations: A Reply to McNeill (1985). In *Psychological Review* (Vol. 94, Issue 4, pp. 493–498). https://doi.org/10.1037/0033-295X.94.4.493

Feyereisen, P. (2007). How do gesture and speech production synchronise? *Current Psychology Letters: Behaviour, Brain and Cognition*, *22*(2), 1–14. https://doi.org/10.4000/cpl.1561

Fiehler, K., Reuschel, J., & Rösler, F. (2009). Early non-visual experience influences proprioceptive-spatial discrimination acuity in adulthood. *Neuropsychologia*, *47*(3), 897–906. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2008.12.023

Fillmore, C. J. (1997). *Lectures on deixis.* Stanford: CSLI Publications.

Finisguerra, A., Canzoneri, E., Serino, A., Pozzo, T., & Bassolino, M. (2015). Moving sounds within the peripersonal space modulate the motor system. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2014.09.043

Forker, D. (2019). Elevation as a category of grammar: Sanzhi Dargwa and beyond . *Linguistic Typology*, *23*(1), 59-106. https://doi.org/10.1515/lingty-2019-0001

Forker, D. (2020). *A grammar of Sanzhi Dargwa (Volume 2)*. Language Science Press.

Freedman, N. (1970). Hands, words, and mind: on the structuralization of body movements during discourse and the capacity for verbal representation. *The Psychoanalytic Quarterly*, *39*(1), 17–37. https://doi.org/10.1080/21674086.1970.11926514

Fricke, E. (2014). 136. Deixis, gesture, and embodiment from a linguistic point of view. In *Volume 2* (pp. 1803-1823). De Gruyter Mouton.

Friesen, C.K., Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5,* 490–495.

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809–1829. https://doi.org/10.1111/brv.12535

Galfano, G., Dalmaso, M., Marzoli, D., Pavan, G., Coricelli, C., & Castelli, L. (2012). Eye gaze cannot be ignored (but neither can arrows). *The Quarterly Journal of Experimental Psychology*, *65*(10), 1895-1910.

Galiano, A. R., & Portalier, S. (2011). Language and Visual Impairment: Literature Review. *International Psychology, Practice and Research,(2)*.

Gallivan, J. P., Cavina-Pratesi, C., & Culham, J. C. (2009). Is that within reach? fMRI reveals that the human superior parieto-occipital cortex encodes objects reachable by the hand. *Journal of Neuroscience*, *29*(14), 4381-4391.

Gallivan, J. P., McLean, A., & Culham, J. C. (2011). Neuroimaging reveals enhanced activation in a reach-selective brain area for objects located within participants' typical hand workspaces. *Neuropsychologia*, *49*(13), 3710-3721.

Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Retrieved from https://gamlj.github.io/.

García, J. O. P., Ehlers, K. R., & Tylén, K. (2017). Bodily constraints contributing to multimodal referentiality in humans: The contribution of a de-pigmented sclera to proto declaratives. *Language & Communication*, *54*, 73-81.

Gaunet, F., & Rossetti, Y. (2006). Effects of visual deprivation on space representation: Immediate and delayed pointing toward memorised proprioceptive targets. *Perception*, *35*(1), 107–124. https://doi.org/10.1068/p5333

Gentilucci, M., & Dalla Volta, R. (2008). Spoken language and arm gestures are controlled by the same motor control system. *Quarterly Journal of Experimental Psychology*, *61*(6), 944–957. https://doi.org/10.1080/17470210701625683

Goodridge, W. H., & Shaheen, N. L., & Hunt, A. T., & & Kane, D (2021). Work in Progress: The Development of a Tactile Spatial Ability Instrument for Assessing Spatial Ability in Blind and Low-vision Populations. *2021 ASEE Virtual Annual Conference Content Access*. Retrieved from https://par.nsf.gov/biblio/10333032.

Goldin-Meadow, S. (2014). How gesture works to change our minds. In *Trends in Neuroscience and Education* (Vol. 3, Issue 1, pp. 4–6). Elsevier GmbH.

Gonseth, C., Vilain, A., & Vilain, C. (2013). An experimental study of speech/gesture interactions and distance encoding. *Speech Communication*, *55*(4), 553–571. https://doi.org/10.1016/j.specom.2012.11.003

González-Peña, P., Doherty, M. J., & Guijarro-Fuentes, P. (2020). Acquisition of Demonstratives in English and Spanish. *Frontiers in Psychology*, *11*(July), 1–13. https://doi.org/10.3389/fpsyg.2020.01778

Gori, M., Bertonati, G., Campus, C., & Amadeo, M. B. (2022). Multisensory representations of space and time in sensory cortices. *Human Brain Mapping*, *44*(2), 656–667. https://doi.org/10.1002/HBM.26090

Gori, M., Campus, C., Signorini, S., Rivara, E., & Bremner, A. J. (2021). Multisensory spatial perception in visually impaired infants. *Current Biology*, *31*(22), 5093-5101.e5. https://doi.org/10.1016/j.cub.2021.09.011

Gori, M., Sandini, G., Martinoli, C., Burr, D. C., & Burr, D. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *A JOURNAL OF NEUROLOGY*, *137*, 288–293. https://doi.org/10.1093/brain/awt311

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, *10*, 57–67

Grassmann, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, *13*(1), 252–263. https://doi.org/10.1111/J.1467-7687.2009.00871.X

Gredebäck, G., Melinder, A., & Daum, M. (2010). The development and neural basis of pointing comprehension. *Social Neuroscience*, *5*(5), 441–450. https://doi.org/10.1080/17470910903523327

Gudde, H. B., Coventry, K. R., & Engelhardt, P. E. (2015). The Influence of Language on Memory for Object Location. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society, CogSci 2015*, 830–835.

Gudde, H. B., Griffiths, D., & Coventry, K. R. (2018). The (Spatial) memory game: Testing the relationship between spatial language, object knowledge, and spatial cognition. *Journal of Visualized Experiments*, *2018*(132).

Gugo, I., & Geld, R. (2017). *Frequency and type of gesture in the blind and the sighted in L1 and L2*. *August*. https://doi.org/10.17234/uzrt.2016.1

Guijarro-Fuentes, P., Gudde, H. B., González-Peña, P., & Coventry, K. R. (2022). Children's use of demonstrative words: spatial deictics beyond infancy. *Journal of Child Language*, 1–13. https://doi.org/10.1017/s030500092200054x

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience*, *23*(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Hanks, William F. (1992). The indexical ground of deictic reference. In Alessandro Duranti & Charles Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon*, (pp. 43– 76). Cambridge: Cambridge University Press.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615. https://doi.org/10.1016/j.jml.2007.01.008

Hanna, J. E., Brennan, S. E., & Savietta, K. J. (2020). Eye Gaze and Head Orientation Cues in Face-to-Face Referential Communication. *Discourse Processes*, *57*(3), 201–223. https://doi.org/10.1080/0163853X.2019.1675467

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*, *216*, 116561.

He, Y., Gebhardt, H., Steines, M., Sammer, G., Kircher, T., Nagels, A., & Straube, B. (2015). The EEG and fMRI signatures of neural integration: An investigation of meaningful gestures and corresponding speech. *Neuropsychologia*, *72*, 27–42. https://doi.org/10.1016/j.neuropsychologia.2015.04.018

Heller, M. A., & Kennedy, J. M. (1990). Perspective taking, pictures, and the blind. *Perception & Psychophysics*, *48*(5), 459–466. https://doi.org/10.3758/BF03211590

Hellwig, B. (2003). The Grammatical Coding of Postural Semantics in Goemai (a West Chadic Language of Nigeria). Nijmegen: Max Planck Series in Psycholinguistics.

Herrmann, S. (2018). Warao demonstratives. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Eds.), *Demonstratives in Cross-Linguistic Perspective* (pp. 282–302) Cambridge: Cambridge University Press.

Hietanen, J. K. (1999). Does your gaze direction and head orientation shift my visual attention?. *Neuroreport*, *10*(16), 3443-3447.

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. In *Trends in Cognitive Sciences* (Vol. 23, Issue 8, pp. 639–652). Elsevier Ltd. https://doi.org/10.1016/j.tics.2019.05.006

Holler, J., Kelly, S.D., Hagoort, P., & Özyürek, A. (2012). When gestures catch the eye: The influence of gaze direction on co-speech gesture comprehension in triadic communication. *Annual Meeting of the Cognitive Science Society*.

Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin and Review*, *25*(5), 1900–1908. https://doi.org/10.3758/s13423-017-1363-z

Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S. D., & Özyürek, A. (2015). Eye'm talking to you: Speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Social Cognitive and Affective Neuroscience*, *10*(2), 255–261. https://doi.org/10.1093/scan/nsu047

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692–697. https://doi.org/10.1016/j.cognition.2014.08.008

Hunley, S. B., & Lourenco, S. F. (2018). What is peripersonal space? An examination of unresolved empirical issues and emerging findings. *Wiley interdisciplinary reviews: Cognitive science*, *9*(6), e1472.

Huo, H., Liu, X., Tang, Z., Dong, Y., Zhao, D., Chen, D., Tang, M., Qiao, X., Du, X., Guo, J., Wang, J., & Fan, Y. (2023). Interhemispheric multisensory perception and Bayesian causal inference. *IScience*, *26*(5), 106706. https://doi.org/10.1016/j.isci.2023.106706

Hutchinson, R. S., & Eardley, A. F. (2020). The Accessible Museum: Towards an Understanding of International Audio Description Practices in Museums. *Journal of Visual Impairment and Blindness*, *114*(6), 475–487. https://doi.org/10.1177/0145482X20971958

Hyslop, C. (1993). *Towards a Typology of Spatial Deixis*. Bachelor's thesis, Australian National University, Canberra

Iachini, T., Ruggiero, G., & Ruotolo, F. (2014). Does blindness affect egocentric and allocentric frames of reference in small and large scale spaces? *Behavioural Brain Research*, *273*, 73–81. https://doi.org/10.1016/J.BBR.2014.07.032

Ionin, T., Baek, S., Kim, E., Ko, H., & Wexler, K. (2012). That's not so different from the: Definite and demonstrative descriptions in second language acquisition. In *Second Language Research*,*28* (1). https://doi.org/10.1177/0267658311432200

Iossifova, R., & Marmolejo-Ramos, F. (2013). When the body is time: Spatial and temporal deixis in children with visual impairments and sighted children. *Research in Developmental Disabilities*, *34*(7), 2173–2184. https://doi.org/10.1016/j.ridd.2013.03.030

Iverson, J. M., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology*, *33*(3), 453–467.

Iverson, J. M., & Goldin-Meadow, S. (2001). The resilience of gesture in talk: Gesture in blind speakers and listeners. *Developmental Science*, *4*(4), 416–422.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness studies*, *6*(11-12), 19-40.

Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive development*, *9*(1), 23-43.

Iverson, J. M., Tencer, H. L., Lany, J., & Goldin-Meadow, S. (2000). The relation between gesture and speech in congenitally blind and sighted language-learners. *Journal of nonverbal behavior*, *24*, 105-130.

Jarbou, S. O. (2010). Accessibility vs. physical proximity: An analysis of exophoric demonstrative practice in Spoken Jordanian Arabic. *Journal of Pragmatics*, *42*(11), 3078–3097. https://doi.org/10.1016/j.pragma.2010.04.014

Jones, B. (1975). Spatial perception in the blind. *British Journal of Psychology*, *66*(4), 461–472. https://doi.org/10.1111/J.2044-8295.1975.TB01481.X

Jones, L., Murray, M., & Gomes, R. S. M. (2022). Moving Towards Inclusivity: A Call for Increased Speed and Intensity in Making Fitness Facilities Accessible for People with Visual Impairments. *Journal of Visual Impairment and Blindness*, *116*(5), 752–755. https://doi.org/10.1177/0145482X221133938

Jungbluth, K. (2003). Deictics in the conversational dyad: Findings in Spanish and some cross-linguistic outlines. In F. Lenz (Ed.), Deictic conceptualisation of space, time and person (pp. 13–40). John Benjamins Publishing Company, Pragmatics & Beyond New Series

Kane, S. K., Wobbrock, J. O., & Ladner, R. E. (2011). Usable gestures for blind people: Understanding preference and performance. *Conference on Human Factors in Computing Systems - Proceedings*, 413–422. https://doi.org/10.1145/1978942.1979001

Kappers, A. M. L. (2003). Large systematic deviations in a bimanual parallelity task: Further analysis of contributing factors. *Acta Psychologica*, *114*(2), 131–145. https://doi.org/10.1016/S0001-6918(03)00063-5

Kawachi, K., Bellingham, E., & Bohnemeyer, J. (2018). Different types of causality and clause linkage in English, Japanese, Sidaama, and Yucatec Maya. In *18th National Conference of the Japanese Cognitive Linguistics Association* (pp. 47-59).

Kelly, S. D., & Lee, A. L. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes, 27*(6), 793-807.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267. https://doi.org/10.1177/0956797609357327

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic bulletin & review*, *22*, 517-523.

Kemmerer, D. (1999). "Near" and "far" in language and perception. *Cognition*, *73*(1), 35–63. https://doi.org/10.1016/S0010-0277(99)00040-2

Kemp, N. J. (1981). Social psychological aspects of blindness: A review. *Current Psychological Reviews*, *1*(1), 69–89. https://doi.org/10.1007/BF02979255

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*(C), 22–63. https://doi.org/10.1016/0001-6918(67)90005-4

Kendon, A. (1975). Gesticulation, Speech, and the Gesture Theory of Language Origins. *Sign Language Studies*, *1009*(1), 349–373. https://doi.org/10.1353/sls.1975.0016

Kendon, A. (1980). Gesticulation and Speech: Two Aspects of the Process of Utterance. In *The Relationship of Verbal and Nonverbal Communication* (pp. 207-228). Berlin, New York: De Gruyter Mouton

Kendon, A. (1994). Do gestures communicate? A review. *Research on language and social interaction*, *27*(3), 175-200.

Kendon, A. (1997). Gesture. *Annual review of anthropology*, *26*(1), 109-128.

Kendon, A. (2004). *Gesture: visible action as utterance.* Cambridge, UK: Cambridge University Press.

Kendon, A. (2011). Gesticulation and Speech: Two aspects of the process of utterance. *The Relationship of Verbal and Nonverbal Communication*, *March*, 207–227. https://doi.org/10.1515/9783110813098.207

Kendon, A., & Cook, M. (1969). The consistency of gaze patterns in social interaction. *British Journal of Psychology*, *60*(4), 481–494. https://doi.org/10.1111/j.2044-8295.1969.tb01222.x

Kennedy, J. M., Gabias, P., & Heller, M. A. (1992). Space, haptics and the blind. *Geoforum*, *23*(2), 175–189. https://doi.org/10.1016/0016-7185(92)90015-V

Khusro, S., Shah, B., Khan, I., & Rahman, S. (2022). Haptic Feedback to Assist Blind People in Indoor Environment Using Vibration Patterns. *Sensors 2022, Vol. 22, Page 361*, *22*(1), 361. https://doi.org/10.3390/S22010361

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Psychology Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Klatzky, R. L., & Lederman, S. J. (2011). Haptic object perception: spatial dimensionality and relation to vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1581), 3097–3105. https://doi.org/10.1098/RSTB.2011.0153

Klatzky, R. L., Lederman, S. J., & Matula, D. E. (1993). Haptic Exploration in the Presence of Vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 726–743.

Kok, K., Bergmann, K., Cienki, A., & Kopp, S. (2016). Mapping out the multifunctionality of speakers' gestures. *Gesture*, *15*(1), 37–59. https://doi.org/10.1075/gest.15.1.02kok

Kolarik, A. J., Moore, B. C. J., Zahorik, P., Cirstea, S., & Pardhan, S. (2015). Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics 2015 78:2*, *78*(2), 373–395. https://doi.org/10.3758/S13414-015-1015-1

Kolarik, A. J., Pardhan, S., Cirstea, S., & Moore, B. C. J. (2017). Auditory spatial representations of the world are compressed in blind humans. *Exp Brain Res*, *235*, 597–606.

Kolarik, A. J., Raman, R., Moore, B. C. J., Cirstea, S., Gopalakrishnan, S., & Pardhan, S. (2020). The accuracy of auditory spatial judgments in the visually impaired is dependent on sound source distance. *Scientific Reports 2020 10:1*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-64306-8

Kowadlo, G., Ye, P. J., & Zukerman, I. (2010). Influence of gestural salience on the interpretation of spoken requests. In T. Kobayashi, K. Hirose, & S. Nakamura (Eds.), *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (pp. 2034 - 2037). Causal Productions.

Krstovska-Guerrero, I., & Jones, E. A. (2015). Social-Communication Intervention for Toddlers with Autism Spectrum Disorder: Eye Gaze in the Context of Requesting and Joint Attention. *Journal of Developmental and Physical Disabilities 2015 28:2*, *28*(2), 289–316. https://doi.org/10.1007/S10882-015-9466-9

Küntay, A. C., & Özyürek, A. (2006). Learning to use demonstratives in conversation: What do language specific strategies in Turkish reveal? *Journal of Child Language*, *33*(2), 303–320. https://doi.org/10.1017/S0305000906007380

Kunz, A., Miesenberger, K., Mühlhäuser, M., Alavi, A., Pölzer, S., Pöll, D., ... & Schnelle-Walka, D. (2014). Accessibility of brainstorming sessions for blind people. In *Computers Helping People with Special Needs: 14th International Conference, ICCHP 2014, Paris, France, July 9-11, 2014, Proceedings, Part I 14* (pp. 237-244). Springer International Publishing.

Làdavas, E. (2002). Functional and dynamic properties of visual peripersonal space. *Trends in cognitive sciences*, *6*(1), 17-22.

Làdavas, E., & Serino, A. (2008). Action-dependent plasticity in peripersonal space representations. *Cognitive Neuropsychology*, *25*(7–8), 1099–1113.

Landau, B. (1991). Spatial representation of objects in the young blind child. *Cognition*, *38*(2), 145–178. https://doi.org/10.1016/0010-0277(91)90050-E

Landau, B., Spelke, E., & Gleitman, H. (1984). Spatial knowledge in a young blind child. *Cognition*, *16*(3), 225–260. https://doi.org/10.1016/0010-0277(84)90029-5

Langton, S. R. H., & Bruce, V. (2000). You must see the point: Automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 747–757. https://doi.org/10.1037/0096-1523.26.2.747

Langton, S. R. H., O'Malley, C., & Bruce, V. (1996). Actions Speak No Louder Than Words: Symmetrical Cross-Modal Interference Effects in the Processing of Verbal and Gestural Information. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1357–1375. https://doi.org/10.1037/0096-1523.22.6.1357

Langton, S. R., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual cognition*, *6*(5), 541-567.

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. In *Trends in Cognitive Sciences* (Vol. 4, Issue 2, pp. 50–59). Elsevier Current Trends. https://doi.org/10.1016/S1364-6613(99)01436-9

Latini Corazzini, L., Tinti, C., Schmidt, S., Mirandola, C., & Cornoldi, C. (2010). Developing spatial knowledge in the absence of vision: allocentric and egocentric representations generated by blind people when supported by auditory cues. *Psychologica Belgica*, *50*, 327-334.

Lederman, S. J. (1996). Cognitive salience of haptic object properties: Role of modality-encoding bias. *Perception*, *25*(8), 983–998.

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology*, *19*(3), 342–368.

Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. Attention, Perception, & Psychophysics 2009 71:7, 71(7), 1439–1459. https://doi.org/10.3758/APP.71.7.1439

Lederman, S. J., Klatzky, R. L., Chataway, C., & Summers, C. D. (1990). Visual mediation and the haptic recognition of two-dimensional pictures of common objects. *Perception & Psychophysics*, *47*(1), 54–64.

Legrand, D., Brozzoli, C., Rossetti, Y., & Farnè, A. (2007). Close to me: Multisensory space representations for action and pre-reflexive consciousness of oneself-in-the-world. *Consciousness and cognition*, *16*(3), 687-699.

Lepore, E., & Ludwig, K. (2000). The semantics and pragmatics of complex demonstratives. *Mind*, *109*(434), 199–240. https://doi.org/10.1093/mind/109.434.199

Levinson, S. C. & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B*, *369,* 20130302.

Levinson, S. C. (1983). *Pragmatics*. Cambridge university press.

Levinson, S. C. (2010). Questions and responses in Yélî Dnye, the Papuan language of Rossel Island. *Journal of Pragmatics, 42*(10), 2741-2755. doi: 10.1016/j.pragma.2010.04.009.

Levinson, S. C. (2018). Demonstratives – Patterns of diversity. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Eds), *Demonstratives in Cross-Linguistic Perspective* (pp.1–42). Cambridge: Cambridge University Press

Levinson, S. C. (2018). Yélî Dnye: demonstratives in the language of Rossel Island, Papua New Guinea. In S. C. Levinson, S. Cutfield, M. J. Dunn, N. J. Enfield, and S. Meira (Eds.) *Demonstratives in Cross-Linguistic Perspective* (pp 318–342). Cambridge: Cambridge University Press. doi: 10.1017/9781108333818.016

Levinson, S. C. (2018). Introduction: Demonstratives: Patterns in Diversity. *Demonstratives in Cross-Linguistic Perspective*, 1–42. https://doi.org/10.1017/9781108333818.002

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651). https://doi.org/10.1098/rstb.2013.0302

Levinson, S. C., Cutfield, S., Dunn, M. J., Enfield, N. J., & Meira, S. (Eds.). (2018). *Demonstratives in cross-linguistic perspective* (Vol. 14). Cambridge University Press.

Levinson, S. (1983). *Pragmatics*. Cambridge University Press, Cambridge

Liszkowski, U. (2010). Deictic and other gestures in infancy. *Acción psicológica*, *7*(2), 21-33.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A Prelinguistic Gestural Universal of Human Communication. *Cognitive Science*, *36*(4), 698–713. https://doi.org/10.1111/J.1551-6709.2011.01228.X

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2011). *A Prelinguistic Gestural Universal of Human Communication*. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, *7*(3), 297–307. https://doi.org/10.1111/J.1467-7687.2004.00349.X

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, *7*(2), 173–187. https://doi.org/10.1207/s15327647jcd0702_2

Longo, M. R., & Lourenco, S. F. (2006). On the nature of near space: Effects of tool use and the transition to far space. *Neuropsychologia*, *44*(6), 977–981. https://doi.org/10.1016/j.neuropsychologia.2005.09.003

Longo, M. R., Musil, J. J., & Haggard, P. (2012). Visuo-tactile integration in personal space. *Journal of Cognitive Neuroscience*, *24*(3), 543-552.

Lopez, S. E., Goodridge, W. H., Gougler, I., Kane, D., & Shaheen, N. L. (2020, April). Preliminary Validation of a Spatial Ability Instrument for the Blind and Low Vision. In *AERA Annual Meeting*.

Lücking, A., Pfeiffer, T., & Rieser, H. (2015). Pointing and reference reconsidered. *Journal of Pragmatics*, *77*, 56-79.

Lyons, J. (1977). *Semantics* (1–2). New York, NY: Cambridge University Press

National Eye Institute (June, 2023). *Visual impairment, blindness cases in U.S. expected to double by 2050*. https://www.nei.nih.gov/about/news-and-events/news/visual-impairment-blindness-cases-us-expected-double-2050

Macaluso, E., & Maravita, A. (2010). The representation of space near the body through touch and vision. *Neuropsychologia*, *48*(3), 782–795. https://doi.org/10.1016/j.neuropsychologia.2009.10.010

Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4), 6–6. https://doi.org/10.1167/13.4.6

Maes, A., Krahmer, E., & Peeters, D. (2022). Understanding demonstrative reference in text: A new taxonomy based on a new corpus. *Language and Cognition*, *14*(2), 185–207. https://doi.org/10.1017/langcog.2021.28

Magosso, E., Ursino, M., di Pellegrino, G., Làdavas, E., & Serino, A. (2010). Neural bases of peri-hand space plasticity through tool-use: Insights from a combined computational-experimental approach. *Neuropsychologia*, *48*(3), 812–830. https://doi.org/10.1016/j.neuropsychologia.2009.09.037

Makin, T. R., Holmes, N. P., & Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. *Journal of Neuroscience, 27*(4), 731-740.

Mamus, E., Speed, L. J., Rissman, L., Majid, A., & Özyürek, A. (2023). Lack of Visual Experience Affects Multimodal Language Production: Evidence From Congenitally Blind and Sighted People. *Cognitive Science*, *47*(1).

Mannella, S. (2021). *Seeing without sight: an exploration of the athlete--guide partnership in high-performance parasport* (Doctoral dissertation, University of British Columbia).

Maravita, A., Spence, C., & Driver, J. (2003). Multisensory integration and the body schema: Close to hand and within reach. *Current Biology*, *13*(13), 531–539. https://doi.org/10.1016/S0960-9822(03)00449-4

Markostamou, I., & Coventry, K. R. (2022). Naming spatial relations across the adult lifespan: At the crossroads of language and perception. *Neuropsychology*, *36*(3), 216.

Marschark, M., & Mayer, T. S. (2014). Mental representation and memory in deaf adults and children. In *Psychological perspectives on deafness* (pp. 53-77). Psychology Press.

Martolini, C., Cappagli, G., Saligari, E., Gori, M., & Signorini, S. (2021). Allocentric spatial perception through vision and touch in sighted and blind children. *Journal of Experimental Child Psychology*, *210*, 105195. https://doi.org/10.1016/J.JECP.2021.105195

McNeill, D. (1985). So You Think Gestures Are Nonverbal? *Psychological Review*, *92*(3), 350–371. https://doi.org/10.1037/0033-295X.92.3.350

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* University of Chicago Press.

McNeill, D. (2006). Gesture: A psycholinguistic approach. In *The Encyclopedia of Language and Linguistics*, 1-15

McNeill, D. (2019). *Gesture and thought*. University of Chicago press.

Meira, S. (2018). Tiriyó – Non-contrastive exophoric uses of demonstratives. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, and S. Meira (Eds), *Demonstratives in Cross-Linguistic Perspective* (pp. 222–241). Cambridge: Cambridge University Press.

Meira, S., & Terrill, A. (2005). Contrasting contrastive demonstratives in Tiriyó and Lavukaleve. *Linguistics*, *43*(6), 1131–1152. https://doi.org/10.1515/ling.2005.43.6.1131

Microsoft Research (2023). *Project Triton- Immersive sound propagation for games and mixed reality*. https://www.microsoft.com/en-us/research/project/project-triton/

Mills, A. E. (1983) Acquisition of speech sounds in the visually-handicapped child. In A. E. Mills (ed.), *Language acquisition in the blind child: normal and deficient*. London: Croom Helm

Monfared, N., & Estaji, A. (2011). A Survey and Description of Deictic terms Usage and Understanding in the language of blind Persian speakers. *Journal of Language Researc)*, *3*(5), 1-24.

Morimoto, T. (2020). The Nature of Haptic Working Memory Capacity and Its Relation to Visual Working Memory. *Multisensory Research*, *33*(8), 837–864. https://doi.org/10.1163/22134808-BJA10007

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, *9*(5), 454–464. https://doi.org/10.1111/j.1467-7687.2006.00512.x

Newport, R., Rabb, B., & Jackson, S. R. (2002). Noninformative Vision Improves Haptic Spatial Perception. *Current Biology*, *12*(19), 1661–1664.

Noel, J. P., Samad, M., Doxon, A., Clark, J., Keller, S., & Di Luca, M. (2018). Peri-personal space as a prior in coupling visual and proprioceptive signals. *Scientific Reports 2018 8:1*, *8*(1), 1–15. https://doi.org/10.1038/s41598-018-33961-3

Norris, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: A multimodal interaction analysis. *Visual Communication*, *10*(2), 129–147.

Obermeier, C., Kelly, S. D., & Gunter, T. C. (2014). A speaker's gesture style can affect language comprehension: ERP evidence from gesture-speech integration. *Social Cognitive and Affective Neuroscience*, *10*(9), 1236–1243. https://doi.org/10.1093/scan/nsv011

Osaba, M., Doro, J., Liberal, M., Lagunas, J., Kuo, I. C., & Reviglio, V. E. (2019). Relationship Between Legal Blindness and Depression. *Medical Hypothesis, Discovery & Innovation Ophthalmology Journal*, *8*(4), 306–311. http://www.ncbi.nlm.nih.gov/pubmed/31788493%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6778679

Ottink, L., Hoogendonk, M., Doeller, C. F., Van der Geest, T. M., & Van Wezel, R. J. A. (2021). Cognitive map formation through haptic and visual exploration of tactile city-like maps. *Scientific Reports 2021,11*(1), 1–12.

Özer, D., Karadöller, D. Z., Özyürek, A., & Göksun, T. (2023). Gestures Cued by Demonstratives in Speech Guide Listeners' Visual Attention During Spatial Language

Comprehension. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001402

Özyürek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face- to-face conversational interaction. In S. Santi, I. Guaitella, C. Cave & G. Konopczynski, (Eds.), *Oralité et gestualité: Communication multimodale, interaction* (pp. 609-614). Paris: L'Harmattan.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651). https://doi.org/10.1098/RSTB.2013.0296

Pagel, M., Atkinson, Q. D., Calude, A. S. & Meade, A. (2013). Ultraconserved words point to deep language ancestry across Eurasia. *PNAS*, *110*, 8471–8476.

Papadopoulos, K., & Koustriava, E. (2011). The impact of vision in spatial coding. *Research in Developmental Disabilities*, *32*(6), 2084–2091. https://doi.org/10.1016/j.ridd.2011.07.041

Pasqualotto, A., & Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1179-1187.

Pasqualotto, A., Finucane, C. M., & Newell, F. N. (2005). Visual and haptic representations of scenes are updated with observer movement. *Experimental Brain Research*, *166*(3–4), 481–488. https://doi.org/10.1007/S00221-005-2388-5/FIGURES/3

Pasqualotto, A., Lam, J. S. Y., & Proulx, M. J. (2013). Congenital blindness improves semantic and episodic memory. *Behavioural Brain Research*, *244*, 162–165.

Pavelin Lešić, B. (2010). The metaphorization of practical action and everyday life experience in the words, emblems, and coverbal gestures of spoken language. *Studia Romanica et Anglica*

Zagrabiensia: Revue publiée par les Sections romane, italienne et anglaise de la Faculté des Lettres de l'Université de Zagreb, *55*, 223-233.

Pearson, A., Ropar, D., & de Hamilton, A. F. C. (2013). A review of visual perspective taking in autism spectrum disorder. *Frontiers in Human Neuroscience*, *7*, 1–10. https://doi.org/10.3389/fnhum.2013.00652

Peeters, D., & Özyürek, A. (2016). This and That Revisited: A Social and Multimodal Approach to Spatial Demonstratives. *Frontiers in Psychology*, *7*, 222. https://doi.org/10.3389/fpsyg.2016.00222

Peeters, D., Azar, Z., & Özyürek, A. (2014). The Interplay between Joint Attention, Physical Proximity, and Pointing Gesture in Demonstrative Choice. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 1144–1149.

Peeters, D., Hagoort, P., & Özyürek, A. (2015). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition*, *136*, 64–84. https://doi.org/10.1016/j.cognition.2014.10.010

Peeters, D., Krahmer, E., & Maes, A. (2021). A conceptual framework for the study of demonstrative reference. *Psychonomic Bulletin & Review*, *28*, 409-433.

Perea García, J. O., Ehlers, K. R., & Tylén, K. (2017). Bodily constraints contributing to multimodal referentiality in humans: The contribution of a de-pigmented sclera to proto-declaratives. *Language & Communication*, *54*, 73–81. https://doi.org/10.1016/J.LANGCOM.2016.10.007

Pérez-Pereira, M. (1999). Deixis, personal reference, and the use of pronouns by blind children. *Journal of Child Language*, *26*(3), 655–680. https://doi.org/10.1017/S0305000999003955

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53-78.

Piwek, P., Beun, R. J., & Cremers, A. (2008). “Proximal” and “distal” in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*(4), 694–718. https://doi.org/10.1016/j.pragma.2007.05.001

Postma, A., Zuidhoek, S., Noordzij, M. L., & Kappers, A. M. L. (2008)a. Haptic orientation perception benefits from visual experience: Evidence from early-blind, late-blind, and sighted people. *Perception and Psychophysics, 70*(7), 1197–1206.

Postma, A., Zuidhoek, S., Noordzij, M. L., & Kappers, A. M. L. (2008)b. Keep an eye on your hands: On the role of visual mechanisms in processing of haptic space. *Cognitive Processing*, *9*(1), 63–68.

Prasov, Z., & Chai, J. Y. (2008). What’s in a Gaze? The role of eye-gaze in reference resolution in multimodal conversational interfaces. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 20–29. https://doi.org/10.1145/1378773.1378777

R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.0) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2021-04-01).

Rayson, H., Bonaiuto, J. J., Ferrari, P. F., Chakrabarti, B., & Murray, L. (2019). Building blocks of joint attention: Early sensitivity to having one’s own gaze followed. *Developmental Cognitive Neuroscience*, *37,* 100631. https://doi.org/10.1016/j.dcn.2019.100631

Reile, M., Plado, H., Gudde, H. B., & Coventry, K. R. (2020). Demonstratives as spatial deictics or something more? Evidence from Common Estonian and Võro. *Folia Linguistica*, *54*(1), 167-195.

Ricciardi, E., Menicagli, D., Leo, A., Costantini, M., Pietrini, P., & Sinigaglia, C. (2017). Peripersonal space representation develops independently from visual experience. *Scientific Reports*, *7*(1), 1–10.

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior*, *5*, 172-183.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194. https://doi.org/10.1016/S0166-2236(98)01260-0

RNIB (June, 2023). *Key information and statistics on sight loss in the UK*. https://www.rnib.org.uk/professionals/health-social-care-education-professionals/knowledge-and-research-hub/key-information-and-statistics-on-sight-loss-in-the-uk/

Roberts C (2002) Demonstratives as definites. In van Deemter K and Kibble R (Eds.), *Information sharing* (pp. 89–136). Stanford, CA: CSLI

Rocca, R., Coventry, K. R., Tylén, K., Staib, M., Lund, T. E., & Wallentin, M. (2020). Language beyond the language system: Dorsal visuospatial pathways support processing of demonstratives and spatial language during naturalistic fast fMRI. *NeuroImage*, *216*, 116128. https://doi.org/10.1016/j.neuroimage.2019.116128

Rocca, R., Tylén, K., & Wallentin, M. (2019). This shoe, that tiger: Semantic properties reflecting manuel affordances opf the referent affect demonsttrative use. ocial modulations of proximal vs. distal space in collaborative interaction. *Scientific Reports, 9*, 14967.

Rocca, R., Wallentin, M., Vesper, C. & Tylén, K. (2019). This is for you: Social modulations of proximal vs. distal space in collaborative interaction. *Scientific Reports, 9*, 14967.

Röder, B., Rösler, F., & Neville, H. J. (2000). Event-related potentials during auditory language processing in congenitally blind and sighted people. *Neuropsychologia*, *38*(11), 1482–1502. https://doi.org/10.1016/S0028-3932(00)00057-9

Rohrer, P.L.; Vilà-Giménez, I.; Florit-Pons, J.; Esteve-Gibert, N.; Ren, A.; Shattuck-Hufnagel, S.; Prieto, P. (2020). The MultiModal MultiDimensional (M3D) Labelling Scheme for the Annotation of Audiovisual Corpora. In *Proceedings of the 7th Gesture and Speech in Interaction (GESPIN*), Stockholm, Sweden.

Rosel, J., Caballer, A., Jara, P., & Oliver, J. C. (2005). Verbalism in the narrative language of children who are blind and sighted. *Journal of Visual Impairment and Blindness*, *99*(7), 413–425.

Rubio-Fernandez, P. (2022). Demonstrative systems: From linguistic typology to social cognition. *Cognitive Psychology*, *139*, 101519. https://doi.org/10.1016/J.COGPSYCH.2022.101519

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (7-55). Academic Press.

Saerberg, S. (2010). "Just go straight ahead "how blind and sighted pedestrians negotiate space. *Senses and Society*, *5*(3), 364–381. https://doi.org/10.2752/174589210X12753842356124

Salomon, R., Noel, J. P., Łukowska, M., Faivre, N., Metzinger, T., Serino, A., & Blanke, O. (2017). Unconscious integration of multisensory bodily inputs in the peripersonal space shapes bodily self-consciousness. *Cognition*, *166*, 174–183. https://doi.org/10.1016/j.cognition.2017.05.028

Schifferstein, H. N. J., Smeets, M. A. M., & Postma, A. (2010). Comparing location memory for 4 sensory modalities. *Chemical Senses*, *35*(2), 135–145. https://doi.org/10.1093/chemse/bjp090

Schmidt, S., Tinti, C., Fantino, M., Mammarella, I. C., & Cornoldi, C. (2013). Spatial representations in blind people: The role of strategies and mobility skills. *Acta Psychologica*, *142*(1), 43–50. https://doi.org/10.1016/J.ACTPSY.2012.11.010

Senft, G. (2004). Aspects of spatial deixis in Kilivila. *Deixis and demonstratives in Oceanic languages* (pp. 59-80). Pacific Linguistics.

Serino, A. (2019). Peripersonal space (PPS) as a multisensory interface between the individual and the environment, defining the space of the self. *Neuroscience and Biobehavioral Reviews*, *99*, 138–159. https://doi.org/10.1016/J.NEUBIOREV.2019.01.016

Serino, A., Canzoneri, E., & Avenanti, A. (2011). Fronto-parietal areas necessary for a multisensory representation of peripersonal space in humans: an rTMS study. *Journal of cognitive neuroscience*, *23*(10), 2956-2967.

Serino, A., Bassolino, M., Farnè, A., & Làdavas, E. (2007). Extended multisensory space in blind cane users. *Psychological Science*, *18*(7), 642–648. https://doi.org/10.1111/j.1467-9280.2007.01952.x

Serino, A., Noel, J. P., Galli, G., Canzoneri, E., Marmaroli, P., Lissek, H., & Blanke, O. (2015). Body part-centered and full body-centered peripersonal space representations. *Scientific Reports*, *5*, 1–14. https://doi.org/10.1038/srep18603

Serino, A., Noel, J. P., Mange, R., Canzoneri, E., Pellencin, E., Ruiz, J. B., Bernasconi, F., Blanke, O., & Herbelin, B. (2017). Peripersonal space: An index of multisensory body-environment interactions in real, virtual, and mixed realities. *Frontiers in ICT*, *4*, 1–12. https://doi.org/10.3389/fict.2017.00031

Sharkey, W. F., Asamoto, P., Tokunaga, C., Haraguchi, G., & McFaddon-Robar, T. (2000). Hand gestures of visually impaired and sighted interactants. *Journal of Visual Impairment & Blindness, 94*(9), 549-563.

Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., & Demuth, K. (2016). Non-referential gestures in adult and child speech: Are they prosodic. In *Proceedings from the 8th international conference on speech prosody* (pp. 836-839). Boston: Boston University.

Skilton, A., & Peeters, D. (2021). Cross-linguistic differences in demonstrative systems: Comparing spatial and non-spatial influences on demonstrative use in Ticuna and Dutch. *Journal of Pragmatics*, *180*, 248–265. https://doi.org/10.1016/j.pragma.2021.05.001

Spence, C., Pavani, F., & Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *Journal of Experimental Psychology: Human Perception & Performance, 26*, 1298-1319.

Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cognitive, Affective, & Behavioral Neuroscience 2004 4:2*, *4*(2), 148–169. https://doi.org/10.3758/CABN.4.2.148

Spence, C., Ranson, J., & Driver, J. (2000). Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention. *Perception & Psychophysics, 62*, 410-424.

Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics*, *26*(1), 31–45. https://doi.org/10.1016/j.jneuroling.2012.02.005

Stevens, J., & Zhang, Y. (2014). Brain mechanisms for processing co-speech gesture: A cross-language study of spatial demonstratives work . *Journal of Neurolinguistics*, *30*, 27–47. https://doi.org/10.1016/j.jneuroling.2014.03.003

Steverson, A., & Crudden, A. (2023). Predictors of Job Satisfaction for People With Visual Impairments. *Journal of Visual Impairment and Blindness*. https://doi.org/10.1177/0145482X231168889

Streeck, J. (1993). Gesture as Communication I: Its Coordination with Gaze and Speech. *Communication Monographs*, *60*(4), 275–299. https://doi.org/10.1080/03637759309376314

Struiksma, M. E., Noordzij, M. L., & Postma, A. (2011). Reference frame preferences in haptics differ for the blind and sighted in the horizontal but not in the vertical plane. *Perception*, *40*(6), 725–738. https://doi.org/10.1068/p6805

Stukenbrock, A. (2015). Deixis in der face-to-face-Interaktion. In *Deixis in der face-to-face-Interaktion*. de Gruyter.

Stukenbrock, A. (2020). Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01779

Talmy, L. (1983). How language structures space. In H. Pick, & L. Acredolo (Eds.). *Spatial orientation: Theory, research and application* (pp. 225–282). New York: Plenum Press.

Talmy, L. (2018). *The targeting system of language*. MIT Press.

Tavakol, K., Dehi, M., Naji, H., & Nasiri, M. (2008). *Original Article Parental anxiety and quality of life in children with blindness in Ababasire institution*. *13*(4), 141–144.

Tenbrink, T., Coventry, K. R., & Andonova, E. (2011). Spatial strategies in the description of complex configurations. *Discourse Processes, 48*(4), 237-266.

ter Bekke, M., Drijvers, L., & Holler, J. (2020). The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech. *Gesture and Speech in Interaction (GESPIN2020), August* 1–6. https://doi.org/10.31234/osf.io/b5zq7

Terenghi, S. (2022). Integrating deictic co-speech gestures in demonstrative forms. *Linguistics in the Netherlands*, *39*, 225–239.

Tfouni, L. V., & Klatzky, R. L. (1983). A discourse analysis of deixis: pragmatic, cognitive and semantic factors in the comprehension of 'this', 'that', 'here' and 'there'. *Journal of Child Language*, *10*(1), 123–133. https://doi.org/10.1017/S0305000900005183

The jamovi project (2021). *jamovi*. (Version 2.2) [Computer Software]. Retrieved from https://www.jamovi.org.

Thoermer, C., & Sodian, B. (2001). Preverbal infants' understanding of referential gestures. *First Language*, *21*, 77-87.

Thurston, M., Thurston, A., & Mcleod, J. (2010). Socio-emotional effects of the transition from sight to blindness. *The British Journal of Visual Impairment*, *28*(2), 90–112. https://doi.org/10.1177/0264619609359304

Tinti, C., Adenzato, M., Tamietto, M., & Cornoldi, C. (2006). Visual experience is not necessary for efficient survey spatial cognition: evidence from blindness. *Quarterly journal of experimental psychology*, *59*(7), 1306-1328.

Todisco, E., Guijarro-Fuentes, P., Collier, J., & Coventry, K. R. (2020) The temporal dynamics of deictic communication. *First Language, 41(2)*, 154-178. https://doi.org/10.1177/0142723720936789

Todisco, E., Rocca, R., & Wallentin, M. (2021). The semantics of spatial demonstratives in Spanish: A Demonstrative Choice Task study. *Language and Cognition*, *13*, 503–533. https://doi.org/10.1017/langcog.2021.11

Tomasello, M. (2003). Constructing a language: A usage-based theory of language acquisition. In *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

Tomasello, M. (2008). *Origins of human communication.* 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*(5), 675-691.

Tosi, A., Pickering, M. J., & Branigan, H. P. (2020). Speakers' use of agency and visual context in spatial descriptions. *Cognition*, *194*, 104070. https://doi.org/10.1016/j.cognition.2019.104070

Tuttle, M., & Carter, E. W. (2022). Examining High-Tech Assistive Technology Use of Students With Visual Impairments. *Journal of Visual Impairment and Blindness*, *116*(4), 473–484. https://doi.org/10.1177/0145482X221120265

van Beers, R. J., Sittig, A. C., & Denier Van Der Gon, J. J. (1996). How humans combine simultaneous proprioceptive and visual position information. *Experimental Brain Research 1996 111:2*, *111*(2), 253–261. https://doi.org/10.1007/BF00227302

van der Stoep, N., Postma, A., & Nijboer, T. C. W. (2017). Multisensory Perception and the Coding of Space. *Neuropsychology of Space: Spatial Functions of the Human Brain*, 123–158. https://doi.org/10.1016/B978-0-12-801638-1.00004-5

van Munster, E. P. J., van der Aa, H. P. A., Verstraten, P., & van Nispen, R. M. A. (2021). Barriers and facilitators to recognize and discuss depression and anxiety experienced by adults with vision impairment or blindness: a qualitative study. *BMC Health Services Research*, *21*(1), 1–10. https://doi.org/10.1186/s12913-021-06682-z

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human brain mapping*, *30*(3), 829-858.

Veerman, L., Heppe, E., Gold, D., & Kef, S. (2019). Intra- and Interpersonal Factors in Adolescence Predicting Loneliness among Young Adults with Visual Impairments. *Journal of Visual Impairment and Blindness*, *113*(1), 7–18. https://doi.org/10.1177/0145482X18818615

Vertegaal, R., Slagter, R., Van der Veer, G., & Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. *Proceedings of the SIGCHI conference on Human factors in computing systems*. 301-308.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130292.

Voss, P. (2016). Auditory spatial perception without vision. *Frontiers in Psychology*, *7*(DEC), 1960. https://doi.org/10.3389/fpsyg.2016.01960

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232. https://doi.org/10.1016/J.SPECOM.2013.09.008

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*(2), 252–258. https://doi.org/10.1007/s00221-004-1899-9

Warren, D. H. (1994). *Blindness and children: An individual differences approach*. New York: Cambridge University Press.

Weiss, P. H., Marshall, J. C., Zilles, K., & Fink, G. R. (2003). Are action and perception in near and far space additive or interactive factors? *NeuroImage*, *18*, 837-846.

Welke, D., & Vessel, E. A. (2022). Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on EEG quality. *NeuroImage*, *256*, 119218. https://doi.org/10.1016/J.NEUROIMAGE.2022.119218

Wilkins, D. (1999). Eliciting contrastive use of demonstratives for objects within close personal space (all objects well within arm's reach). In D. Wilkins (Ed.), *Manual for the 1999 Field Season* (pp. 25-28). Nijmegen: Max Planck Institute for Psycholinguistics. doi:10.17617/2.2573796.

Williams, L. E., Huang, J. Y., & Bargh, J. A. (2009). The scaffolded mind: Higher mental processes are grounded in early experience of the physical world. *European Journal of Social Psychology*, *39*(7), 1257–1267. https://doi.org/10.1002/ejsp.665

Wnuczko, M., & Kennedy, J. M. (2014). Pointing to azimuths and elevations of targets: Blind and blindfolded-sighted. *Perception*, *43*(2–3), 117–128. https://doi.org/10.1068/p7605

Wolter, L. (2006). That's that: The semantics and pragmatics of demonstrative noun phrases. Doctoral Dissertation, University of California Santa Cruz.

Zapata, M. A. (2022). Group Identity in Blindness Groups Predicts Life Satisfaction and Lower Anxiety and Depression. *Rehabilitation Psychology*, *67*(1), 42–52. https://doi.org/10.1037/rep0000432

Zuidhoek, S., Visser, A., Bredero, M. E., & Postma, A. (2004). Multisensory integration mechanisms in haptic space perception. *Experimental Brain Research*, *157*(2), 265–268.

Appendices

Table 6.1 Summary of mixed logit model comparing levels of variables via Helmert contrast in Experiment 1**.**

The first level in contrast is the contrast of proximal target with the average of rest of the levels, distal and/or absent. For pointing and gaze, second level of contrast looks at contrast of distal target to no pointing or gaze.

Predictors (fixed effects)	Parameter estimates				Wald's test	
	Log-odds (β)	S.E.	Odds ratio[95%CI]		Z	*p* (β=0)
Intercept	-0.22	0.12	0.80	[-0.46; 0.02]	-1.83	0.07
Noun	4.89	0.60	133.29	[3.72; 6.06]	8.20	< .001
Pointing1	0.97	0.36	2.65	[0.28;1.67]	2.73	0.01
Pointing 2	-0.33	0.30	0.72	[-0.91;0.25]	-1.10	0.27
Gaze1	0.16	0.14	1.18	[-0.12;0.44]	1.14	0.25
Gaze2	-0.02	0.16	0.98	[-0.34;0.29]	-0.14	0.89
Noun1*Pointing1	1.58	0.35	4.86	[0.89;2.27]	4.48	< .001
Noun1*Pointing2	-1.77	0.36	0.17	[-2.47; -1.07]	-4.98	< .001
Noun1*Gaze1	0.33	0.28	1.40	[-0.22; 0.89]	1.18	0.24
Noun1*Gaze2	-0.56	0.32	0.57	[-1.18; 0.07]	-1.74	0.08
Noun1*Gaze1	0.24	0.32	1.27	[-0.39; 0.87]	0.75	0.45
Pointing1*Gaze2	0.23	0.32	1.26	[-0.40; 0.87]	0.71	0.48
Pointing2*Gaze1	0.05	0.36	1.06	[-0.65;0.75]	0.15	0.88
Pointing2*Gaze2	0.62	0.37	1.86	[-0.11; 1.34]	1.68	0.09
Noun1*Pointing1*Gaze1	-0.20	0.64	0.82	[-1.45; 1.06]	-0.31	0.76
Noun1*Pointing2*Gaze1	-0.50	0.65	0.61	[-1.77; 0.77]	-0.77	0.44
Noun1*Pointing1*Gaze2	1.78	0.71	5.90	[0.38; 3.17]	2.49	0.01
Noun1*Pointing2*Gaze2	-0.22	0.74	0.80	[-1.67; 1.22]	-0.30	0.76

Random components	SD	Variance	ICC
Participant (Intercept)	0.487	0.237	0.07
Noun	3.389	11.486	
Pointing1	1.648	2.716	
Pointing 2	1.283	1.646	

Number of observations 2736 grouped by participants (N= 38)

Table 6.2: Summary of mixed logit model comparing levels of variables via Helmert contrast in Experiment 2.

The first level in contrast is the contrast of proximal target with the average of rest of the levels, distal and/or absent. For pointing and gaze, second level of contrast looks at contrast of distal target to no pointing or gaze.

Predictors (fixed effects)	Parameter estimates				Wald's test	
	Log-odds (β)	S.E.	Odds ratio[95%CI]		Z	*p* (β=0)
Intercept	0.41	0.18	1.51	[1.07; 2.13]	2.37	0.02
Demonstrative1	0.82	0.47	2.28	[0.91; 5.72]	1.76	0.08
Pointing1	5.68	0.65	294.26	[82.14;1054.25]	8.73	< .001
Pointing 2	-3.85	0.53	0.02	[0.01; 0.06]	-7.32	< .001
Gaze1	1.16	0.17	3.19	[2.28; 4.45]	6.80	< .001
Gaze2	-1.51	0.20	0.22	[0.15; 0.33]	-7.43	< .001
Demonstrative1*Pointing1	0.41	0.45	1.51	[0.62; 3.67]	0.90	0.37
Demonstrative1*Pointing2	0.41	0.42	1.51	[0.67; 3.44]	0.99	0.32
Demonstrative1*Gaze1	0.00	0.34	1.00	[0.52; 1.94]	0.01	0.99
Demonstrative1*Gaze2	-0.30	0.40	0.74	[0.34; 1.64]	-0.74	0.46
Pointing1*Gaze1	-1.30	0.40	0.27	[0.12; 0.60]	-3.25	0.00
Pointing1*Gaze2	-0.38	0.36	0.68	[0.34; 1.38]	-1.06	0.29
Pointing2*Gaze1	1.28	0.46	3.61	[1.46; 8.88]	2.79	0.01
Pointing2*Gaze2	1.48	0.45	4.40	[1.81; 10.71]	3.26	0.00
Demonstrative1*Pointing1*Gaze1	-1.20	0.80	0.30	[0.06; 1.45]	-1.50	0.13
Demonstrative1*Pointing2*Gaze1	-0.54	0.72	0.58	[0.14; 2.40]	-0.75	0.46
Demonstrative1*Pointing1*Gaze2	0.73	0.92	2.07	[0.34; 12.53]	0.79	0.43
Demonstrative1*Pointing2*Gaze2	-0.81	0.91	0.44	[0.07; 2.66]	-0.89	0.38

Random components	SD	Variance	ICC
Participant (Intercept)	0.528	0.278	0.22
Demonstrative	2.313	5.35	
Pointing 1	2.851	8.126	
Pointing 2	2.407	5.794	

Number of observations 2178 grouped by participants (N=34).

Table 6.3: Summary of mixed logit model comparing levels of variables via Helmert contrast in Experiment 3.

The first level in contrast is the contrast of proximal target with the average of rest of the levels, distal and/or absent. For pointing and gaze, second level of contrast looks at contrast of distal target to no pointing or gaze.

Predictors (fixed effects)	Parameter estimates				Wald's test	
	Log-odds (β)	S.E.	Odds ratio	[95%CI]	Z	*p* (β=0)
Intercept	0.61	0.12	1.83	[1.46; 2.30]	5.18	< .001
Demonstrative1	1.16	0.36	3.19	[1.57; 6.47]	3.21	0.00
Pointing1	4.00	0.22	54.50	[35.21; 84.37]	17.93	< .001
Pointing 2	-3.50	0.22	0.03	[0.02; 0.05]	-15.77	< .001
Gaze1	1.43	0.19	4.20	[2.87; 6.13]	7.42	< .001
Gaze2	-2.24	0.23	0.11	[0.07; 0.17]	-9.58	< .001
Demonstrative1*Pointing1	-0.75	0.45	0.47	[0.20; 1.13]	-1.68	0.09
Demonstrative1*Pointing2	-1.15	0.44	0.32	[0.13; 0.76]	-2.59	0.01
Demonstrative1*Gaze1	0.13	0.39	1.13	[0.53; 2.42]	0.33	0.75
Demonstrative1*Gaze2	0.38	0.47	1.47	[0.59; 3.67]	0.82	0.41
Pointing1*Gaze1	-2.60	0.43	0.07	[0.03; 0.17]	-6.05	< .001
Pointing1*Gaze2	3.38	0.52	29.45	[10.56; 82.13]	6.46	< .001
Pointing2*Gaze1	-1.69	0.45	0.18	[0.08; 0.45]	-3.76	< .001
Pointing2*Gaze2	3.38	0.52	29.45	[10.56; 82.13]	6.46	< .001
Demonstrative1*Pointing1*Gaze1	0.66	0.86	1.94	[0.36; 10.42]	0.77	0.44
Demonstrative1*Pointing2*Gaze1	-0.30	0.90	0.74	[0.13; 4.29]	-0.34	0.74
Demonstrative1*Pointing1*Gaze2	-0.10	1.06	0.90	[0.11; 7.17]	-0.10	0.92
Demonstrative1*Pointing2*Gaze2	0.03	1.05	1.03	[0.13; 7.99]	0.03	0.98

Random components	SD	Variance	ICC
Participant (Intercept)	0.10	0.61	0.05
Demonstrative	1.77	3.14	

Number of observations 2376, grouped by participants (N=33)

Table 6.4: Summary of mixed logit model comparing levels of variables via Helmert contrast in Experiment 4. For stimuli type, the contrast compares dynamic to static stimuli. The first level in contrast in demonstratives, pointing and gaze are the contrast of proximal target with the average of rest of the levels, distal and/or absent. For pointing and gaze, second level of contrast looks at contrast of distal target to no pointing or *gaze.*

Predictors (fixed effect)	Parameter estimates				Wald's test	
	Log-odds (β)	S.E.	Odds ratio[95%CI]		Z	*p* (β=0)
Intercept	-0.04	0.09	0.96	[0.80;1.15]	-0.43	0.67
Type1	-0.48	0.12	0.62	[0.49; 0.78]	-4.07	< .001
Demonstrative1	1.81	0.23	6.13	[3.90; 9.64]	7.84	< .001
Pointing1	6.29	0.28	539.10	[309.97;937.59]	22.28	< .001
Pointing 2	-4.57	0.26	0.01	[0.01; 0.02]	-17.6	< .001
Gaze1	1.30	0.20	3.67	[2.48; 5.43]	6.51	< .001
Gaze2	-0.23	0.19	0.79	[0.54; 1.16]	-1.21	0.23
Type1*Demonstrative1	-0.24	0.24	0.79	[0.50; 1.25]	-1.02	0.31
Type1*Pointing1	0.22	0.26	1.25	[0.74; 2.09]	0.84	0.40
Type1*Pointing2	-0.86	0.27	0.42	[0.25; 0.72]	-3.17	0.00
Type1*Gaze1	0.85	0.23	2.34	[1.49; 3.69]	3.68	< .001
Type1*Gaze2	-0.62	0.31	0.54	[0.30; 0.98]	-2.01	0.04
Demonstrative1*Pointing1	-0.27	0.29	0.76	[0.43; 1.35]	-0.93	0.35
Demonstrative1*Pointing2	-1.93	0.31	0.15	[0.08; 0.27]	-6.29	< .001
Type1*Demonstrative1*Pointing1	0.82	0.53	2.27	[0.81; 6.36]	1.56	0.12
Type1*Demonstrative1*Pointing2	-1.66	0.54	0.19	[0.07; 0.55]	-3.06	0.00
Demonstrative1*Gaze1	-0.37	0.24	0.69	[0.44; 1.10]	-1.55	0.12
Demonstratice1*Gaze2	-0.15	0.31	0.86	[0.47; 1.58]	-0.50	0.62
Type1*Demonstrative1*Gaze1	0.59	0.46	1.80	[0.73; 4.46]	1.27	0.20
Type1*Demonstrative1*Gaze2	-1.31	0.61	0.27	[0.08; 0.90]	-2.14	0.03
Pointing1*Gaze1	-2.36	0.29	0.09	[0.05; 0.17]	-8.11	< .001
Pointing1*Gaze2	-1.52	0.37	0.22	[0.11; 0.45]	-4.12	< .001
Pointing2*Gaze1	-2.33	0.31	0.10	[0.05; 0.18]	-7.53	< .001
Pointing2*Gaze2	0.62	0.38	1.86	[0.88; 3.92]	1.63	0.10
Type1* Pointing1*Gaze1	-0.91	0.52	0.40	[0.15; 1.11]	-1.75	0.08
Type1* Pointing1*Gaze2	-0.43	0.69	0.65	[0.17; 2.51]	-0.62	0.53

Type1* Pointing2*Gaze1	-0.55	0.53	0.58	[0.20; 1.65]	-1.02	0.31
Type1* Pointing2*Gaze2	1.69	0.71	5.41	[1.35; 21.63]	2.39	0.02
Demonstrative1*Pointing1*Gaze1	1.02	0.55	2.78	[0.95; 8.14]	1.86	0.06
Demonstrative1*Pointing2*Gaze1	-0.49	0.57	0.61	[0.20; 1.85]	-0.87	0.38
Demonstrative1*Pointing1*Gaze2	-0.51	0.72	0.60	[0.15; 2.48]	-0.70	0.48
Demonstrative1*Pointing2*Gaze2	0.47	0.72	1.60	[0.39; 6.54]	0.65	0.52
Type1* Demonstrative1*Pointing1*Gaze1	0.65	1.03	1.91	[0.25; 14.41]	0.63	0.53
Type1* Demonstrative1*Pointing1*Gaze2	-0.45	1.37	0.64	[0.04; 9.35]	-0.33	0.74
Type1* Demonstrative1*Pointing2*Gaze1	-0.72	1.07	0.49	[0.06; 3.95]	-0.67	0.50
Type1* Demonstrative1*Pointing2*Gaze2	0.85	1.41	2.34	[0.15; 37.09]	0.60	0.55

Random Components	SD	Variance	ICC
Intercept	0.41	0.17	.005
Demonstrative1	1.53	2.34	
Pointing1	1.77	3.14	
Pointing2	1.43	2.03	
Gaze1	1.17	1.37	
Gaze2	0.76	0.58	

Number of observations 10512, grouped by participants (N=73)

Table 6.5: Summary of mixed logit model comparing levels of variables via Helmert contrast in Experiment 4, static trials. The first level in contrast is the contrast of proximal target with the average of rest of the levels, distal and/or absent. For pointing and gaze, second level of contrast looks at contrast of distal target to no pointing or gaze.

Predictors (fixed effects)	Parameter estimates				Wald's test	
	Log-odds (β)	S.E.	Odds ratio[95%CI]		Z	*p* (β=0)
Intercept	0.23	0.11	1.26	[1.02; 1.57]	2.12	0.03
Demonstrative1	1.76	0.24	5.81	[3.64; 9.27]	7.39	< .001
Pointing1	5.92	0.33	370.78	[195.59; 702.88]	18.13	< .001
Pointing 2	-3.78	0.25	0.02	[0.01; 0.04]	-15.05	< .001
Gaze1	0.80	0.14	2.22	[1.67; 2.95]	5.53	< .001
Gaze2	-0.09	0.19	0.92	[0.63; 1.33]	-0.46	0.64
Demonstrative1*Pointing1	-0.64	0.40	0.53	[0.24; 1.16]	-1.59	0.11
Demonstrative1*Pointing2	-1.10	0.35	0.33	[0.17; 0.66]	-3.17	0.00
Demonstrative1*Gaze1	-0.65	0.29	0.52	[0.30; 0.92]	-2.26	0.02
Demonstratice1*Gaze2	0.40	0.38	1.49	[0.71; 3.12]	1.05	0.29
Pointing1*Gaze1	-1.91	0.34	0.15	[0.08; 0.29]	-5.61	< .001
Pointing1*Gaze2	-2.01	0.31	0.13	[0.07; 0.24]	-6.52	< .001
Pointing2*Gaze1	-1.67	0.45	0.19	[0.08; 0.46]	-3.67	< .001
Pointing2*Gaze2	-0.08	0.39	0.92	[0.43; 2.00]	-0.20	0.84
Demonstrative1*Pointing1*Gaze1	0.99	0.68	2.70	[0.71; 10.22]	1.46	0.15
Demonstrative1*Pointing2*Gaze1	-0.52	0.62	0.60	[0.18; 2.00]	-0.84	0.40
Demonstrative1*Pointing1*Gaze2	-0.64	0.91	0.53	[0.09; 3.12]	-0.70	0.48
Demonstrative1*Pointing2*Gaze2	0.49	0.79	1.64	[0.35; 7.67]	0.63	0.53

Random components	SD	Variance	ICC
Participant (Intercept)	0.354	0.125	0.04
Demonstrative	1.348	1.817	
Pointing1	1.637	2.68	
Pointing2	1.251	1.565	

Number of observations 5256, grouped by participants (N=73)

Table 6.6: Summary of mixed logit model comparing levels of variables via Helmert contrast in Experiment 4, dynamic trials. The first level in contrast is the contrast of proximal target with the average of rest of the levels, distal and/or absent. For pointing and gaze, second level of contrast looks at contrast of distal target to no pointing or gaze.

Predictors (fixed effects)	Parameter estimates				Wald's test	
	Log-odds (β)	S.E.	Odds ratio[95%CI]		Z	*p* (β=0)
Intercept	-0.26	0.15	0.77	[0.58; 1.02]	-1.81	0.07
Demonstrative1	1.68	0.32	5.39	[2.87; 10.13]	5.24	< .001
Pointing1	6.65	0.41	772.03	[347.10; 1717.18]	16.30	< .001
Pointing 2	-5.23	0.41	0.01	[0.00; 0.01]	-12.66	< .001
Gaze1	1.67	0.18	5.30	[3.71; 7.58]	9.15	< .001
Gaze2	-0.65	0.24	0.52	[0.32; 0.84]	-2.69	0.01
Demonstrative1*Pointing1	0.09	0.44	1.10	[0.46; 2.60]	0.21	0.83
Demonstrative1*Pointing2	-3.01	0.53	0.05	[0.02; 0.14]	-5.66	< .001
Demonstrative1*Gaze1	-0.04	0.36	0.96	[0.47; 1.95]	-0.12	0.90
Demonstratice1*Gaze2	-0.76	0.49	0.47	[0.18; 1.22]	-1.56	0.12
Pointing1*Gaze1	-2.90	0.39	0.06	[0.03; 0.12]	-7.44	< .001
Pointing1*Gaze2	-2.68	0.44	0.07	[0.03; 0.16]	-6.06	< .001
Pointing2*Gaze1	-2.18	0.52	0.11	[0.04; 0.31]	-4.19	< .001
Pointing2*Gaze2	1.61	0.59	4.98	[1.58; 15.76]	2.74	0.01
Demonstrative1*Pointing1*Gaze1	1.60	0.78	4.93	[1.07; 22.63]	2.05	0.04
Demonstrative1*Pointing2*Gaze1	-1.43	0.88	0.24	[0.04; 1.35]	-1.62	0.11
Demonstrative1*Pointing1*Gaze2	-1.07	1.04	0.34	[0.04; 2.63]	-1.03	0.30
Demonstrative1*Pointing2*Gaze2	1.40	1.17	4.04	[0.40; 40.29]	1.19	0.23

Random components	SD	Variance	ICC
Participant (Intercept)	0.37	0.137	0.04
Demonstrative	1.87	3.50	
Pointing1	2.22	4.91	
Pointing 2	1.68	2.81	

Number of observations 5256, grouped by participants (N=73)

Table 6.7: Summary of mixed logit model comparing levels of variables via Helmert contrast in Experiment 5.

Predictors (fixed effects)	Parameter estimates				Wald's test	
	Log-odds (β)	S.E.	Odds ratio[95%CI]		Z	*p* (β=0)
Intercept	-1.57	0.55	0.21		-2.87	0.00
A - (B, C, D, E, F, G, H, I, J, K, L)	2.15	0.19	8.58	[0.07; 0.60]	11.18	< .001
B - (C, D, E, F, G, H, I, J, K, L)	2.15	0.19	8.55	[5.89; 12.51]	11.23	< .001
C - (D, E, F, G, H, I, J, K, L)	2.02	0.19	7.57	[5.88; 12.44]	10.76	< .001
D - (E, F, G, H, I, J, K, L)	1.78	0.19	5.91	[5.24; 10.95]	9.48	< .001
E - (F, G, H, I, J, K, L)	0.64	0.20	1.91	[4.09; 8.53]	3.22	0.00
F - (G, H, I, J, K, L)	0.21	0.22	1.23	[1.29; 2.82]	0.97	0.33
G - (H, I, J, K, L)	0.55	0.21	1.73	[0.81; 1.88]	2.58	0.01
H - (I, J, K, L)	0.17	0.23	1.19	[1.14; 2.63]	0.76	0.45
I - (J, K, L)	0.50	0.24	1.64	[0.76; 1.86]	2.10	0.04
J - (K, L)	0.24	0.26	1.28	[1.03; 2.61]	0.93	0.35
K - L	-0.09	0.31	0.92	[0.77; 2.12]	-0.28	0.78
Combination- (Haptic, Visual)	1.56	1.08	4.78	[0.50; 1.68]	1.44	0.15
Haptic - Visual	0.27	1.43	1.31	[0.57; 39.90]	0.19	0.85
A - (B, C, D, E, F, G, H, I, J, K, L) * Combination- (Haptic, Visual)	0.42	0.41	1.52	[0.08; 21.67]	1.03	0.30
B - (C, D, E, F, G, H, I, J, K, L) * Combination- (Haptic, Visual)	0.19	0.39	1.21	[0.69; 3.35]	0.48	0.63
C - (D, E, F, G, H, I, J, K, L) *Combination- (Haptic, Visual)	0.34	0.39	1.41	[0.56; 2.61]	0.89	0.38
D - (E, F, G, H, I, J, K, L) * Combination- (Haptic, Visual)	0.38	0.38	1.46	[0.66; 3.0]	0.99	0.32
E - (F, G, H, I, J, K, L) * Combination- (Haptic, Visual)	-0.51	0.40	0.60	[0.69; 3.07]	-1.25	0.21
F - (G, H, I, J, K, L) * Combination- (Haptic, Visual)	0.37	0.42	1.45	[0.27; 1.33]	0.88	0.38

G - (H, I, J, K, L) * Combination- (Haptic, Visual)	0.43	0.42	1.54	[0.63; 3.32]	1.03	0.30
H - (I, J, K, L) * Combination- (Haptic, Visual)	0.37	0.45	1.45	[0.68; 3.51]	0.82	0.41
I - (J, K, L) x Combination- (Haptic, Visual)	0.45	0.47	1.57	[0.60; 3.51]	0.96	0.34
J - (K, L) * Combination- (Haptic, Visual)	-0.10	0.53	0.91	[0.63; 3.91]	-0.19	0.85
K - L * Combination- (Haptic, Visual)	-0.05	0.62	0.95	[0.320; 2.53]	-0.08	0.93
A - (B, C, D, E, F, G, H, I, J, K, L) *Haptic - Visual	-0.10	0.47	0.90	[0.28; 3.22]	-0.21	0.83
B - (C, D, E, F, G, H, I, J, K, L) * Haptic - Visual	-0.66	0.48	0.52	[0.36; 2.29]	-1.37	0.17
C - (D, E, F, G, H, I, J, K, L) *Haptic - Visual	-0.41	0.48	0.66	[0.20; 1.33]	-0.86	0.39
D - (E, F, G, H, I, J, K, L) * Haptic - Visual	-0.12	0.48	0.89	[0.26; 1.69]	-0.26	0.80
E - (F, G, H, I, J, K, L) * Haptic - Visual	-0.02	0.51	0.98	[0.35; 2.26]	-0.05	0.96
F - (G, H, I, J, K, L) *Haptic - Visual	0.78	0.56	2.19	[0.36; 2.66]	1.39	0.17
G - (H, I, J, K, L) * Haptic - Visual	0.33	0.56	1.39	[0.72; 6.59]	0.59	0.55
H - (I, J, K, L) * Haptic - Visual	0.77	0.60	2.15	[0.47; 4.16]	1.29	0.20
I - (J, K, L) *Haptic - Visual	-0.58	0.62	0.56	[0.67; 6.91]	-0.94	0.35
J - (K, L) *Haptic - Visual	0.04	0.66	1.04	[0.17; 1.88]	0.06	0.95
K - L * Haptic - Visual	-0.49	0.79	0.61	[0.28; 3.82]	-0.62	0.54

-

Random components	SD	Variance	ICC
Participant (Intercept)	0.22	0.05	0.01
Condition 1	6.06	36.73	
Condition 2	9.39	88.18	

Number of observations 3888, grouped by participants (N=81)