Production and Comprehension of Audience Design Behaviours in Co-Speech Gesture.

By

Jacob Barker

Thesis

Submitted to the University of Warwick

For the degree of

Doctor of Philosophy in Psychology

University of Warwick, Department of Psychology

February 2022

Contents

List of Figures and Tables

Acknowledgements

I would like to thank the Economic and Social Research Council and the Midlands Graduate School for providing me with the funding necessary to undertake the PhD. The support and training they have provided me with have been invaluable in developing my skills as a researcher. I would also like to thank the Department of Psychology at the University of Warwick for hosting my PhD and giving me the opportunity to create this thesis.

I would like to earnestly thank my first supervisor, Professor Sotaro Kita. Professor Kita's support and guidance during my time at the University of Warwick have given me the confidence and skills to conduct and disseminate my research. Professor Kita has always given his full support to both my work, and my development as a researcher. It has been a joy to collaborate on so many projects together, and I sincerely hope we can continue to work together in the future. Furthermore, I would like to thank the other members of the Department of Psychology I have had the pleasure to work with including Dr James Adelman, Professor Derrick Watson, Dr Hester Duffy, Dr Friederike Schlaghecken, and Professor Thomas Hills.

I would like to thank my fellow lab members who have provided so much feedback and support over the years, helping me to develop and improve my research ideas. I would like to mention Dr Suzanne Aussems, Jiahao Yang, Dr Charlotte Hills, and Melissa Reddy, and wish them all the best in their own research careers. Furthermore, I would like to thank all the research assistants who have provided both direct and indirect help to my thesis including Stéphen Theron-Grimaldi, Daniella Waluube, Sijjal Riaz, Zoe Gardler, and Levent Emir Ozder.

I would also like to thank all the administrative staff from the Department of Psychology who have supported me from my time as an undergraduate, all the way to the end of my PhD studies. I would like to mention Steve McGladrigan, Emily Biggs, Linda Wilson, Emily Stilwell, Jody Roche, Catherine Johnstone, Nicola Grant, and Liz Farmer.

I would also like to thank all of the participants who have taken part in my studies over the years. Without their earnest contribution of time and effort, the research in this thesis would not have been possible.

I would also like to sincerely thank my family and friends for their continual support. I would like to thank Meghna Ebenezer, who has helped me constantly through every day of my PhD and encouraged me to produce this thesis despite the COVID pandemic. I would also like to thank my family: Stephen, William, June, and David Barker for supporting and visiting me during my PhD. I would also like to thank Gareth Jones, Hugh Jones, and Matthew Lumb for providing friendship and levity throughout my PhD.

Declaration

This thesis is submitted to the University of Warwick in support of the application for the degree of Doctor of Philosophy. It has been composed by the author and the written work has not been submitted in any previous application for any degree. The data for Experiments 1 and 2 of Chapter 3 were originally collected for previous postgraduate degrees. The data has been recoded, re-analysed, and presented in a new format for this thesis. The paper which makes up Chapter 3 was entirely produced during my PhD. The work presented (including data collection and data analyses) was carried out by the author.

Parts of this dissertation are being prepared for publication by the author:

Chapter 3: Barker, J., & Kita, S. (2021). The Addressee's Face as a Trigger for Gesture Production: Evidence for Cue-Based Heuristics in Audience Design. *Manuscript Submitted for Publication.*

Chapter 4: Barker, J., & Kita, S. (2022). Gesture in Context: Investigating the Factors and Mechanisms that Affect Co-Speech Gesture Production. *Manuscript Submitted for Publication.*

Chapter 5: Barker, J., & Kita, S. (2022). Foreground Gestures: The Effect of Foregrounding Gestures on Addressee Comprehension. *Manuscript Submitted for Publication.*

Summary

Speakers can use gesture to depict information during conversation (Kendon, 2004). The current thesis investigates how speakers can adjust their gestures to communicate more effectively to an addressee using gesture. Furthermore, the current thesis investigates the mechanisms behind audience design behaviours.

Chapter 1 introduces the topics of gestures and audience design, and outlines the structure of the thesis.

Chapter 2 explores the definition and classification of gestures, and provides a review of the literature on gesture production, gesture comprehension, and audience design.

Chapter 3 investigates the mechanisms responsible for producing audience design behaviours, and the competing factors affecting gesture production. The findings suggest that speakers use cue-based heuristics to design communicative behaviours. Furthermore, the findings suggest that speakers value gesture more for communication when describing spatial stimuli than abstract stimuli.

Chapter 4 further investigates the mechanisms responsible for producing audience design behaviours and the factors affecting gesture production. The findings suggest that speakers can both respond to cues from the addressee using heuristics and take the perspective of the addressee. Furthermore, we found no evidence to suggest that the effect of visibility was due to the confounding of visibility and addressee responsiveness.

Chapter 5 investigates how foregrounding gestures can help the gestures convey information to the addressee. The findings do not provide unequivocal evidence that foregrounding benefits the addressee's comprehension. However, trends in the data suggest that making gestures visually prominent or referring to the gesture in speech may help the gesture to convey information to the addressee.

Chapter 6 discussed and interpreted the findings from the previous Chapters. It discusses the mechanisms responsible for audience design behaviours, the factors that affect gesture production, and the effect of gestural audience design behaviours on addressee comprehension. The chapter discusses my interpretations of the findings regarding the current literature and proposes further research.

Chapter 1

General Introduction

1.1 The use of Gestures to communicate

When we speak, we often produce gestures as well. Speakers produce hand and arm movements when they talk (McNeill, 1992). Co-speech gestures are spontaneously produced during speech and are typically co-ordinated with the words in timing and meaning (McNeill, 1992). Hand gestures can depict phenomena (Clark, 2016; Streeck, 2008) and act as a symbol (McNeill, 1992). Co-speech gestures are thought to be produced as part of the same process used for language (McNeill, 1985).

Speakers can produce different types of gestures. Representational gestures can be iconic, metaphorical, or deictic (pointing). Speakers can produce "iconic" gestures by forming the gestures to depict an action, movement, object, shape or location (McNeill, 1992). Speakers can produce metaphorical gestures to depict an abstract concept, such as weighing up two options while making a descision (McNeill, 1992). Speakers can point to not only physical objects, but also previously established locations in the space around them (McNeill, 1992, 2003). Interactive gestures support the conversation with the addressee and can directly reference the addressee or the interaction itself (Bavelas et al., 1992, 2008). For instance, a speaker could cite the previous contributions of the addressee, seek a response from the addressee, deliver new information, and signify whose turn it is to speak (Bavelas et al., 1995).

Speakers can use gesture to depict information during conversation (Kendon, 2004). Gestures can convey information relevant to speech to help an addressee understand the speaker's message (Hostetter, 2011; Riseborough, 1981). Gestures can also convey additional information not expressed in speech, providing new information to the addressee (Broaders & Goldin-Meadow, 2010). Gestures can provide a supplementary source of information if the addressee doesn't comprehend the message (Sueyoshi & Hardison, 2005). Furthermore, gestures are especially effective at conveying spatial information (Alibali, 2005).

Gestures can successfully convey information to the addressee. Speakers communicate more effectively when producing gestures than when gestures are prohibited (Hostetter, 2011). Furthermore, gestures that convey information which was not present in speech are more effective at conveying information to the addressee than gestures which convey information which is redundant with speech (Hostetter, 2011). Gestures are more effective at conveying information when they depict motoric information such as actions, or spatial dynamic information such as the movement of objects than when depicting abstract metaphorical information (Hostetter, 2011). Gestures are also more effective when the speaker is conveying information to addressees who are less verbally proficient, such as children (Hostetter, 2011). Gestures may help ground abstract ideas for addressees who struggle to understand the verbal message (Hostetter, 2011). Overall, gestures provide a moderate benefit to the addressee's understanding, which is affected by the type of information being conveyed by the gestures.

1.2 Communicative factors that affect speaker's gesture production

Speakers produce gesture differently depending on the interactional context. More specifically, speakers can produce gesture differently depending on whether the addressee can see the gestures, whether the addressee provides responses and feedback to the speaker, and whether the information being conveyed is dynamic, spatial, or abstract.

Speakers produce more gestures when the addressee can see the speaker's gestures because the gestures can convey information. Speakers produce more gestures overall when speaking to an addressee who can see the speaker's gestures rather than an addressee who cannot see the speaker's gestures (Emmorey & Casey, 2001; Hostetter & Potthoff, 2012; Krauss et al., 1995; Mol et al., 2009a, 2009b). Furthermore, speakers produce more representational or illustrator gestures (Alibali et al., 2001; Cohen, 1977; Cohen & Harrison, 1973) and more interactive gestures (Bavelas et al., 1992, 2008) when speaking to an addressee who can see the speaker's gestures. Speakers may produce more gestures to a visible addressee than a non-visible addressee because the gestures can only convey information to a visible addressee. When the addressee cannot see the speaker's gestures, then speakers are not motivated to communicate with gesture. Furthermore, speakers produce larger representational gestures when the addressee can see the speaker's gestures (Bavelas et al., 2008). Speakers also placed gestures higher up in space when the addressee can see the speaker's gestures (Holler et al., 2011). Speakers may produce gestures they wish to help communicate higher up in front of them and make them bigger. The larger, higher up gestures may be clearer and help the speaker convey information. Taken together, the findings suggest that when the addressee can see the speaker's gestures, then speakers are motivated to communicate with gesture.

Speakers produce more gestures when conversing with an addressee who provides responses and feedback. Speakers produce more co-speech gestures when talking to responsive addressees than non-responsive addressees (Beattie & Aboudan, 1994). Speakers and addressees can provide information back and forth between them to ensure the communication is effective (Schober & Clark, 1989). When the speaker communicates a piece of information, the addressee can signal their understanding and the speaker can confirm the understanding is correct (Yngve, 1970). Addressees can also provide non-verbal responses (Bavelas et al., 2011; Bavelas & Gerwing, 2011). Gaze creates a back-channel response through which messages can be sent (Bavelas et al., 2002). Furthermore, addressees can use facial expressions to respond to the speaker. These include smiles (Brunner, 1979) and mimicking the speaker's facial expressions (Bavelas, 2007; Bavelas et al., 2000). When addressees provide feedback to the speaker, speakers produce more gestures in response to the addressee to confirm the message is understood correctly (Bavelas et al., 2011). Speakers also produce fewer gestures when communicating with a less attentive addressee (Jacobs & Garnham, 2007). Taken together, the findings suggest that when addressees provide responses to the speaker, then speakers are motivated to produce more gestures.

Speakers produce more gestures when discussing spatial stimuli than abstract stimuli. Speech and co-speech gesture can be elicited using motoric stimuli, dynamic spatial stimuli, static spatial stimuli, and abstract stimuli. Motor stimuli refers to actions. Dynamic spatial stimuli refers to information such as the path an actor takes relative to objects. Static spatial stimuli refers to information such as the shape of an object. Abstract information includes descriptions of conceptual phenomena such as thoughts, feelings, and opinions. When gestures convey spatial or motoric information, gesture becomes more effective at conveying information (Hostetter, 2011). Gestures are particularly suited to conveying spatial information rather than abstract information because gestures can efficiently convey action, shape, relative location of objects and the movement of objects (Hostetter, 2011). Speakers produce gestures more when describing spatial information than non-spatial information (Alibali, 2005; Rauscher et al., 1996) and addressees comprehended the speaker's message better when seeing gestures conveying spatial and motoric gestures, but not when seeing gestures conveying abstract information (Hostetter, 2011).

Several factors that affect gesture production may have been confounded with each other. The effect of visibility on representational gestures can potentially be explained by the effect of responsiveness (see Bavelas & Healing, 2013, for a review). When speakers and addressees cannot see each other, then both the speaker's gestures and the addressee's visual responses to the speaker are also blocked. Thus, whether the speaker's gestures are visible or not is confounded with whether that addressee's responses are visible or not. When the addressee's visual responses are hidden from speakers, then speakers may produce fewer gestures. The effect of responsiveness on speaker's gesture production could account for the finding that speakers produce fewer gestures when behind a screen (e.g. Alibali et al., 2001).

1.3 Designing gestures for communication

Speakers can design their communicative efforts to effectively convey a message to an addressee. Audience design refers to the ways in which speakers tailor their communication to benefit their addressee (Clark & Murphy, 1982). Speakers can design their gestures to convey a message more effectively to the addressee. For example, speakers will orient their gestures to face towards the space they share with their addressee when speaking to a single addressee, or within a central space when speaking to two addressees (Özyürek, 2002).

The effect of visibility on gesture is an audience design effect. Speakers produce more gestures when an addressee can see the speaker's gestures (e.g. Alibali et al., 2001). As mentioned previously, the effect of visibility on gesture has been attributed to the communicative function of gestures. When the addressee see the speaker's gestures, then speakers are motivated to communicate with gesture. However, the explaination above assumes that speakers are sensitive to whether their gestures can be seen by the addressee. Speakers must take the perspective of the addressee and determine that the addressee can see the gestures.

It is possible that speakers produce gestures in response to seeing the cues from the addressee, rather than taking the perspective of the addressee. When the speakers' behaviour seems to be tailored for their addressee, it does not necessarily mean that speakers designed those behaviours for the addressee (Barr & Keysar, 2006). Speakers can instead take advantage of cues available in their interaction with the addressee (Shintel & Keysar, 2009). Such cues can trigger speakers to change their behaviour in pre-determined ways. The addressee's face being visible could be such a cue. When speakers see the addressee's face, a heuristic process could be triggered and increase the rate of gesture production.

The addressee's face may be a cue for triggering gesture production because being able to see each other's faces is an important part of conversation. Addressees typically fixate constantly on the speaker's face during conversation (Gullberg & Holmqvist, 2006). Speakers are sensitive to the gaze of the addressee and use speech and gesture to draw the addressee's gaze towards the speaker (C. Goodwin, 1986; M. H. Goodwin & Goodwin, 1986). Addressees use facial expressions to provide feedback to the speaker, which can impact the speaker's communication (Bavelas et al., 2000; Bavelas & Gerwing, 2011).

There is little empirical evidence to unequivocally support a cue-based mechanism of audience design behaviours such as the effect of visibility on gesture. To our knowledge no study has a design in which cues from the addressee have triggered an audience design behaviour that cannot be explained by the speaker taking the perspective of the addressee. Most support for the cue-based heuristic mechanism relies on the argument that heuristics are a cognitively economical alternative to the perspective taking mechanism (Galati & Brennan, 2010, 2014; Shintel & Keysar, 2009). Furthermore, no study has separated mutual gesture visibility from mutual facial visibility. Without separating gesture visibility and facial visibility, we cannot attribute the effect of visibility on gesture to speaker's inferring their gestures can be seen and being motivated to communicate with gesture. It is possible that when speakers see the addressee's face, a heuristic process could be triggered and increase the rate of gesture production instead.

1.4 The effect of gestural audience design behaviours on communicative effectiveness

Speakers can place gestures in the foreground of an interaction to convey information more effectively. Speakers can design their utterances to attract the addressee's attention and clearly convey important information (Cooperrider, 2018). Speakers can refer to their gesture in speech using demonstratives (Guérin, 2015). Demonstratives can include phrases such as 'like this' or 'like that'. By referring to the gesture in speech, gesture becomes the focus of communication. Speakers can also fixate their gaze on their own gestures (Gullberg & Kita, 2009). Speaker can fixate on their own gestures to place the gestures in the foreground of the interaction. Addressees are sensitive to the speaker's gaze fixation on gesture (Gullberg & Kita, 2009). Speakers can also put noticeable effort into gesture production to place them in the foreground of the interaction. Speakers can make their gestures larger (Bavelas et al., 2008) or more precise (Gerwing & Bavelas, 2004). By putting more effort into the gestures, speakers are signalling the importance of the gesture for conveying the message to the addressee.

Placing gestures in the foreground of the interaction may draw the addressee's attention to the gestures. Addressees typically fixate constantly on the speaker's face during conversation rather than the speaker's gestures (Gullberg & Holmqvist, 2006). Speakers are also sensitive to the gaze of the addressee and use speech and gesture to draw the addressee's gaze towards the speaker (C. Goodwin, 1986; M. H. Goodwin & Goodwin, 1986). By placing gestures in the 'foreground' of an interaction, the speaker can potentially draw the addressee's attention to the speaker's gesture. For instance, speakers can fixate their gaze on their own gestures to place them in the foreground of the interaction. Addressees fixate on the speaker's gesture more when the speaker fixates on the gestures (Gullberg & Kita, 2009).

It has not been well established that placing gestures in the foreground affects the addressee's comprehension. No prior study has provided unequivocal evidence that placing gestures in the foreground leads to improved addressee comprehension. Gestures provide an overall benefit to addressee comprehension, especially when conveying spatial and non-redundant information (Hostetter, 2011). There is little evidence to suggest that audience design behaviours such as using larger gestures benefits the addressee's comprehension. Speakers fixating on their gestures benefit addressee's comprehension of directional information (whether a character goes left or right) (Gullberg & Kita, 2009). However, only an effect on direction information was tested, and the effect was not found when the artificial (but controlled) gaze fixations were added to the stimuli. Gesture duration may have been confounded with gaze fixation as gestures were longer when the speaker fixated on the gestures that when the speaker did not fixate on the gestures. No study has investigated the effect of speakers producing larger gestures on addressee comprehension. Furthermore, no study has investigated the effect of speakers producing gestures indicated by demonstratives on addressee comprehension. If speakers do design gestures to communicate more effectively with the addressee, then addressees should benefit from the audience design behaviours.

1.5 Thesis Overview

In the current thesis, we have investigated three questions important to the use of gesture in communication. First, we have investigated whether the effect of visibility on gesture can be explained by a cue-based mechanism by which speakers are cued to produce more gestures upon seeing the addressee's face. Second, we have investigated which factors can affect co-speech gesture including gesture visibility, face visibility, addressee responsiveness and stimuli topic. Third, we have investigated if gestural audience design behaviours result in addressees better comprehending the speaker's message.

We have investigated whether the effect of visibility on gesture production can be explained by cue-based mechanisms. A cue-based mechanism has been put forward as a plausible mechanism but has not been empirically supported. Furthermore, the traditional explanation of the effect of visibility on gesture that assumes speakers take the perspective of the addressee to infer if gestures are visible. The perspective taking explanation has not previously been compared to the cue-based explanation.

We have investigated the factors affecting co-speech gestures to establish how speakers adjust their gestures to convey information more effectively to the addressee depending on the interactional context. Several contextual factors have potentially been confounded in prior studies including mutual visibility of gesture, mutual visibility of faces, the addressee's responsive behaviour (both visual and verbal), and the topic of discussion. It is important to resolve the different explanations for the effect of visibility on gesture as researchers may come to erroneous conclusions if the findings could be explained by several different factors.

We have investigated whether foregrounding gestures helps the addressee comprehend the speaker's message. Several gestural audience design behaviours have been documented such as putting more effort into gesture, fixating gaze on gesture, and using demonstratives to highlight gesture. The behaviours have been used to explain speakers trying to communicate more effectively with addressees and are often used as dependant variables to indicate speakers are trying to communicate (e.g. Bavelas et al., 2008). It is important to investigate the impact of gestural audience design behaviours to establish that behaviours attributed to gestural communication.

In Chapter 2, we have investigated the literature regarding gesture research to help answer the aims of this thesis. First, we examined the literature on the definition and function of gestures first to establish the phenomena we have investigated. We examined both the self-oriented and communication function of gestures to understand why people gesture, and in what situations gestures are useful. For the self-oriented function of gesture, we examined when speakers produce gestures for their own benefit, and how the gestures affect the speaker. For the communicative function of gesture, we have examined evidence for when speakers produce gestures to communicate, as well as evidence that gestures benefit the addressee's comprehension of the speaker's message. We have also investigated the issue of when the speaker intends a gesture to communicate. We have then examined the audience design literature broadly to establish how speakers design communication to convey information effectively to the addressee, before examining how audience design research can apply to gestural behaviours as well.

In Chapter 3, we have conducted a series of experiments to investigate the mechanisms behind audience design effects and how speakers use gestures to communicate when discussing abstract or dynamic spatial topics. We investigated whether speakers took the perspective of the addressee and produced more gestures when gestures were visible to the addressee, or if speakers responded to cues from the addressee and produced more gestures when the addressee's face was visible. Furthermore, we investigated how speakers produced gestures to communicate both when describing abstract stimuli, and when describing dynamic spatial stimuli.

In Chapter 4, we have investigate further competing factors that affect gesture production, and the mechanisms behind audience design effects. Speakers discussed stimuli with either a conversational or silent participant addressee. We investigated if speakers took the perspective of the addressee and produced more gestures when gestures were visible to the addressee, or if speakers responded to cues from the addressee, thus producing more gestures when the addressee's face was visible. We also investigated if the effect of mutual visibility on gesture production is caused by the confounding of gesture visibility and addressee responses.

In Chapter 5, we have investigated the effects of audience design behaviours on addressee comprehension. Over three experiments, participants watched videos of a speaker describing cartoons, and drew target events from the cartoons. In each experiment, the speaker produced foreground gestures either by producing larger gestures, fixated their gaze on gestures, or producing gestures indicated by demonstratives. We investigated the effect of the speaker's audience design behaviours on addressee's memory of target information encoded only in gesture.

In Chapter 6, we have discussed and interpreted the findings from the previous Chapters. We discuss the mechanisms responsible for audience design behaviours, the factors that affect gesture production, and the effect of gestural audience design behaviours on addressee comprehension. We have discussed my interpretations of the findings regarding the current literature and propose further research.

Chapter 2

Literature Review

2.1 Defining and categorising gestures

We can define gesture as the hand and arm movements people spontaneously make when speaking and thinking. The movement and arrangement of the arms and hands in a way that embodies meaning can be called gesture (McNeill, 1992). Gestures are typically produced alongside speech during an interaction (co-speech gestures) but can also be produced when the speaker is silent and alone (co-thought gestures). Gestures are not tied to a specific form but can be "free and reveal the idiosyncratic imagery of thought" (McNeill, 1992). Gesture production is thought to be linked to both our thought and speech processes (Kendon, 2004; McNeill, 1992, 2005).

Gestures are made up of several phases; the preparation, stroke, hold, and retraction phases (Kita et al., 1997; McNeill, 1992). The preparation phase moves from a 'rest' position (such as on the speaker's lap) to where the stroke phase will begin. The stroke phase is the effortful movement of the gesture in which meaning can potentially be found. All 'completed' gestures consist of at least a stroke phase. The hold phase can be before and after the stroke. The arms and hands are held still either in anticipation of the stroke or in the position the stroke ended. The retraction phase is the return of the hands back to a resting position. The stroke phase is considered obligatory for a movement to be considered a gesture and is typically where gesture researchers look for the meaning and category of the gesture.

Gestures have been broken down into different categories. Ekman and Friesen (1969) categorise gestures based on the usage and origin of the gesture. Emblems are gestures which have an established definition and are used in an intentional attempt to communicate. One example is the OK gesture, where the gesturer touches the tip of the index finger and thumb to form a circle while the other fingers are extended upwards. Illustrators are classed as movements that are tied to speech. Ekman and Friesen (1969) break illustrators down into six categories. Batons (commonly known as beats) are bi-phasic movements to place emphasis on a word or phrase. Ideographs are movements which display a path or direction of thought. Deictic gestures are pointing gestures. Spatial movements depict spatial relationships. Kinetographs depict bodily actions. Finally, pictographs depict a picture in the air in the shape of a referent. Apart from emblems and illustrators, Ekman and Friesen (1969) also define regulators, gestures which maintain the conversation between the speaker and addressees. Regulators may be used to manage turn taking or ask for further information on a topic.

McNeill (1992) developed and refined previous gesture category schemes to form four classes of gestures. Iconic gestures depict in their form and manner an aspect of the meaning being presented in speech. Iconic gestures can refer to actions, movements, objects, shapes, or locations. For example, if a speaker says "The cat climbed the ladder" while placing one hand and above the other in sequence, the gesture can be interpreted as the hand movements of pulling oneself up each rung on a ladder. Metaphorical gestures can refer to metaphorical concepts or properties. For example, if a speaker says, "I don't know which party I'd go to" while raising and lowering their hands, alternating which hand is higher, the gesture can be interpreted as the speaker weighing up the two choices as if each choice existed on a different hand and each choice had "weight". Deictic gestures once again refer to pointing gestures. Beat gestures (batons) again refer to bi-phasic movements to place emphasis on a word or phrase.

Bavelas et al. (1992) altered previous gesture categories to include a new group of gestures known as interactive gestures. The category of interactive gestures includes Ekman and Friesen's (1969) regulators, as well as some illustrator gestures. Illustrator gestures (Ekman & Friesen, 1969) can be broken down into topic or interactive gestures based on the subject of the gesture. Topic gestures refer to the topic of discourse (such as iconic or metaphorical gestures), while interactive gestures refer to an aspect of conversing with the addressee. Interactive gestures can cite the previous contributions of the addressee, seeking a response from the addressee such as asking for help, delivering new information, and signifying whose turn it is to speak (Bavelas et al., 1995). An example of speakers citing the previous contributions of the addressee might be pointing a hand towards the addressee in the equivalent of saying "as you said earlier". An example of speakers seeking help from the addressee might be the speaker rotating their hand while searching for a word. An example of speakers delivering new information to the addressee might be the speaker presenting an outstretched palm-upwards hand towards the addressee, as if the information is being physically given to the addressee. An example of speakers signifying who's turn it is to speaker might be the speaker sweeping their hand to the side, with an open empty hand to signify the speaker has nothing more to say and the addressee can take a turn in the conversation. Gestures previously categorised as beat gestures were included as interactive gestures (Bavelas et al., 1992).

Chu et al. (2014) created a scheme for categorising gestures taking into account previous categorisations (Bavelas et al., 1992; McNeill, 1992) which were adapted for use in the current project. Representational gestures (Kita, 2000) are made up of depictive, deictic and conduit gestures. Depictive gestures are movements which can depict the actions of an agent, the movement of an object or a property of either an actor or agent. These are made up of two sub-categories; Iconic and metaphorical gestures (McNeill, 1992). Deictic gestures are pointing gestures which identify a location in the space (McNeill et al., 1993). Interactive gestures (Bavelas et al., 1992) are made up of listener indexing, conduit and palm revealing emblem gestures. Listener-Indexing gestures refer to the addressee. A gesture is coded as a listener-indexing gesture when the hand moving towards the addressee references the addressee. Conduit gestures (Bavelas et al., 1992; McNeill, 1992) are produced by the speaker moving an open palm towards the addressee to present an idea. The speaker should be certain about what they are saying. Conduit gestures can be identified by a clear forward movement of the hand toward the addressee. Conduit gestures are considered both interactive gestures as well as representational gestures. This is because conduit gestures directly index the addressee and represent the presentation of an idea on the speaker's palm. Palm revealing gestures included shrugs with the palms out and facing upwards to express empty handedness.

2.2 Theories of Gesture

Gestures have been of interest to the research community because of the role of gesture in embodied cognition. The theory of embodied cognition states that the mind does not process information in isolation. Instead, the state of the body, sensory and motor systems, and the world around us all contribute to how information is processed (Barsalou, 2008; Foglia & Wilson, 2013; Wilson, 2002). Gesture is thought to both be an expression of how we think (McNeill, 1992, 2005), and influences the way we think (Kita et al., 2017). For example, speakers can point to spaces which have been established to represent something such as the "good guys" or the "bad guys" in a movie (McNeill, 1992). Speakers can produce representational gestures which represent actions, shapes, or objects, which demonstrates the mental simulations of actions (Alibali & Nathan, 2012; Hostetter & Alibali, 2008, 2019; McNeill, 1992). Speakers can also produce gestures conveying metaphorical concepts (McNeill, 1992). Metaphors are often based on space or movement such as good being up and bad being down (Lakoff & Johnson, 1980). Metaphorical gestures demonstrate the embodiment of conceptual metaphors (Alibali & Nathan, 2012).

Gesture production is linked to speech production processes. Gestures are produced in time with speech, leading researchers to suggest that the processes responsible for speech production are also involved in the gesture production process (Goldin-Meadow, 1999; McNeill, 1992). One theory for the link between speech and gesture production is the free imagery hypothesis. Speech and gesture may both be generated at the conceptual stage of language production (de Ruiter, 2000). According to Levelt's model of speech production the conceptualiser collects and prepares information ready to be expressed (Levelt, 1989). The free imagery hypothesis suggests that both speech and gesture are formed from the output (pre-verbal message) of the conceptualiser (de Ruiter, 2000). According to the lexical semantic hypothesis, gestures are generated from the semantic features (particularly spatial ones) from lexical items (Butterworth & Hadar, 1989). The interface hypothesis suggests that gestures are produced not only from the speech production process, but the interfacing of the process that generates practical actions with the speech production process (Kita & Özyürek, 2003). The spatial-motoric representation of information that is prepared for speaking is also used to generate gestures. Furthermore, both gestures produced alone during silent thought (co-thought gestures) and co-speech gestures are produced by the same action generation process (Chu & Kita, 2016). Speakers produced more co-thought and co-speech gestures when thinking about objects which are easily interacted with (e.g., a smooth mug) than objects that would be difficult to interact with (e.g., a spikey mug). Taken together, the findings suggest that gestures are a result of both the speech and action planning processes.

Gestures may be the by-product of action production process. According to the "Gesture as Simulated Action (GSA) Framework", gestures result from simulations of action (Hostetter & Alibali, 2008, 2019). When speakers think about visuospatial or motor imagery, then perceptual and action processes are activated. The perceptual and action processes occurring without external input are referred to as motoric and perceptual simulations. The simulations have associated motor plans, which can be expressed in gesture (thinking about a cup leads to activation of the action of holding or drinking from a cup etc). The simulations activate the motor system. When simulations are highly tied to actions, greater activity is elicited in the motor system. When the motor system is engaged for use in the speech production process, gestures become a likely output of the motor system activity. If the activation of the motor system surpasses a speaker's resistance to producing a gesture (known as the gesture threshold), then a gesture is produced. Speakers may manipulate their gesture threshold on a variety of factors. When speakers are reluctant to gesture, they will raise the threshold of activation required for motor activity to result in a gesture (though sometime motoric activations may be so high that reduced versions of the gestures are still produced (Kita & Essegbey, 2001). When speakers are keen to produce gestures, they will lower the threshold of activation required for motor activity to result in a gesture. The gestures produced reflect the motor plan associated with the simulations.

The many theories of gesture can be divided into two groups, the mechanistic and functional theories of gesture (Özer & Göksun, 2020). Mechanistic theories are concerned with how we gesture. As discussed previously, mechanistic theories suggest we produce gestures as an expression of our cognitive processes. Gestures may arise from speech product process (e.g. Butterworth & Hadar, 1989; de Ruiter, 2000), the simulation or representation of visual, motoric, and action based imagery (e.g. Hostetter & Alibali, 2008, 2019), or the interfacing of speech production and action generation systems (e.g. Kita & Özyürek, 2003). Functionalist theories are instead concerned with why people produce gestures. Speakers are thought to produce gestures to serve both self-oriented and communicative functions. In the next section I will examine the literature investigating the self-oriented function of gesture.

2.3 Gesturing for ourselves: The self-oriented function of gesture

Speakers can produce gestures to aid cognitive processes involving information (Kita et al., 2017). According to the gesture-for-conceptualisation hypothesis, speakers produce gestures to activate, manipulate, package, and explore spatio-motoric information (Kita et al., 2017). Gestures can activate perceptual-motor representations (Hostetter & Boncoddo, 2017) and maintain spatial imagery to prevent the representation from decaying (Wesp et al., 2001). Gestures can help manipulate spatial-motoric information such as rotating three dimensional mental representations of objects (Chu & Kita, 2008, 2011). Gestures can also help speakers to package information into processing units ready for speaking (Alibali et al., 2000). If speakers are encouraged to produce gestures that encompass two pieces of information (e.g. manner and path), then the speakers are more likely to refer to the two pieces of information in a single clause (Mol & Kita, 2012). If speakers are encouraged to produce two pieces of information in separate gestures, then speakers are more likely to refer to the two pieces of information in separate clauses (Mol & Kita, 2012). Finally, gesture can help us explore for potential solutions for a problem (e.g. Broaders et al., 2007). Research into the self-oriented function of gesture comes in two categories: studies measuring the production of gesture and studies measuring gesturer’s performance on tasks.

2.3.1 Producing self-oriented gestures in response to difficult tasks

Speakers produce more gestures when the cognitive demands of a task are difficult. For example, speakers produce more gestures when describing images from memory than when describing images that were visible at the time of description (Morsella & Krauss, 2004; Wesp et al., 2001). The findings suggest that when trying to maintain an image, gesture is used as a tool to prevent the representation decaying in working memory.

Speakers produce more gestures when being placed under a high conceptual load (Kita & Davies, 2009; Melinger & Kita, 2007). Conceptual load was manipulated by increasing the complexity of pictures which participants described, introducing a secondary task, or adding a competing conceptual representation. The findings suggest that under high conceptual load, gesture is used as a tool to lighten the cognitive load (Goldin-Meadow et al., 2001) when components of working memory are overtaxed. Furthermore, the more complex pictures may have been more difficult to describe because packaging the information into units was more difficult. Speakers produced comparable utterances in the easy and difficult conditions, suggesting the increased rate of gesturing in the difficult condition benefited speakers' information packaging. Having a verbal secondary task only marginally increased speaker's gesture rates, while a motoric secondary task which greatly increased speaker's gesture rates (Hoetjes & Masson-Carro, 2017). Thus, gestures may help speakers under high conceptual load to focus in spatio-motoric information (Alibali et al., 2011) which may be especially useful when multiple spatial or motoric representations are competing for attention. Further research suggests that under high conceptual load during a spatial memory task, speakers produce a higher rate of complex iconic-deictic gestures (representing both an object/action and a location in space) (Suppes et al., 2015). However, speakers may have produced a higher rate of complex gestures to communicate more effectively rather than to offset cognitive load as the gestures were produced in a potentially communicative context (Suppes et al., 2015). Taken together, the findings suggest that people use gestures in situations where a task is difficult, and gesture is suited to help. However, studies where gesture production is being investigated can provide limited empirical support for the self-oriented benefit of producing the additional gestures. Researchers can instead investigate the effect of gesture production on task performance.

2.3.2 Producing self-oriented gestures benefits task performance

Speakers perform better at certain tasks when they are allowed to gesture than when they are prohibited from gesturing. For example, participants completed more mental block rotation tasks when gesture was encouraged than when gesture was prohibited (Chu & Kita, 2011; Emmorey & Casey, 2001). Participants were shown three images of objects made from blocks and asked which of two objects the target image matched (Chu & Kita, 2011). Gesture may have benefitted participants by allowing the manipulation of spatial imagery. The gestures helped participants to manipulate the visualisation of the target object in thought. Speakers also benefit from being able to manipulate the visualisation of target objects in a communicative task (Emmorey & Casey, 2001).

When discussing spatial content, speakers are more fluent when gesture is allowed rather than prohibited (Rauscher et al., 1996). When discussing non-spatial content, speakers were equally fluent when allowed or prohibited from gesturing (Rauscher et al., 1996). The findings suggest that gesture is particularly beneficial for helping speakers activate motoric representations, which can active words associated with those representations (Kita et al., 2017). Furthermore, speakers who can gesture use richer verbs (Hostetter et al., 2007), more vivid imagery (Rimé et al., 1984), are more likely to retrieve words on the tip of their tongues (Pine et al., 2007), and are more likely to use spatial language (Emmorey & Casey, 2001). However, the finding of gestures benefitting fluency has not always been found. When describing how to tie a tie, a very motoric subject, speakers were not less fluent or more monotonous when prohibited from gesturing (Hoetjes et al., 2014).

Gestures reduce the cognitive load during demanding tasks (Goldin-Meadow et al., 2001). As discussed previously, speakers produce more gestures under high conceptual or cognitive load (Hoetjes & Masson-Carro, 2017; Kita & Davies, 2009; Melinger & Kita, 2007; Suppes et al., 2015). The benefit of the additional gestures has been explored by prohibiting gesture or allowing it. Both adults and children remembered more items from a list while explaining how to solve a mathematics problem (Goldin-Meadow et al., 2001). The findings suggested that the cognitive load of explaining the maths problem was reduced, allowing participants to allocate cognitive resources to remembering the list.

2.3.3 Individual differences in self-oriented gesture production

Gesturers can benefit a greater or lesser amount from producing gestures depending on the task and individual differences between gesturers (Özer & Göksun, 2020). For instance, people with lower visual and spatial working memory capacities, spatial transformation ability, and conceptualization ability use more gestures than higher visual-spatial ability people (Chu et al., 2014). The finding suggests that participants who would struggle to maintain the scenario in working memory use gesture to help maintain the imagery. Furthermore, the findings suggest that participants who would struggle with spatial transformation used more gesture to help manipulate visio-spatial imagery. Similarly to how participants produce more gestures when spatial transformation is difficult (e.g. Chu & Kita, 2011), participants produce more gestures when poorer spatial transformation and conceptualisation abilities make the task more difficult for them.

Speakers with lower verbal working memory and semantic fluency abilities produced more gestures while speaking than speakers with higher verbal abilities (Hostetter & Alibali, 2007; Smithson & Nicoladis, 2013). The findings suggest that gestures may be used to compensate for weaker verbal abilities. Gestures could be used to help the speaker's produce more fluent utterances (Rauscher et al., 1996) by helping the speaker to package information (Kita et al., 2017). Furthermore, individuals who had weaker verbal skills but stronger spatial skills produced a higher rate of gestures (Hostetter & Alibali, 2007, 2011), suggesting gestures are useful when trying to convey visio-spatial information preferred by high spatial ability speakers, and producing the images in gesture can help turn the spatial knowledge into verbal utterances (Hostetter & Alibali, 2007). Speakers with strong spatial skills but weaker verbal skills also produce a higher proportion of non-redundant gestures (Gestures which convey information which is not present in speech). The finding suggests speakers with strong spatial skills and weaker verbal ones may avoid putting some information into verbal utterances by conveying the information in gesture.

Taken all together, the findings suggest that speakers produce gestures for self-oriented purposes. The literature suggests both that gestures helps individuals activate, manipulate, package, and explore spatio-motoric information, and that individuals respond to difficult tasks by producing more gestures.

2.4 **Gesturing for others: The communicative function of gesture**

Speakers can produce gestures to convey information to addressees (Kendon, 1994, 2004). Kendon, (2004) breaks down the contribution of gestures into six categories. First, gestures with specific pre-determined meanings (emblems) which are produced alongside speech conveying the same information. For example, speakers could produce a 'thumbs-up" emblem while saying they approve of a choice an addressee has made. Second, emblems which do not overlap with speech and contain non-redundant meaning. For example, speakers could produce an 'OK' emblem without any speech to let the addressee know that all is well. These first two ways in which gestures contribute information rely on pre-existing definitions of the gestures produced. Third, gestures can be produced alongside speech to add specificity to the verbal utterance. For example, when talking about someone climbing up a wall, the gesture can add information to the verb 'climb' by representing the climbing up of a ladder, rope, or handholds in the wall. Fourth, gestures can represent objects. Speakers can represent objects by shaping their hands to form the object. For example, a speaker could represent the hands on a clock with their fingers. Speakers can also represent objects by using their hands to form around the object. For instance, a speaker could represent a basketball by placing the hands palms-opposed to each other and curving the hand in a slight semi-circle shape as if each is holding one side of the ball. Fifth, gestures can represent spatial information such as shape, size, and relative location. For example, the hands can be spaced a specific distance apart to demonstrate the width of an object, or the shape of an object can be specified by tracing the outline in the air. Finally, speakers can establish locations in space using gesture. For example, a speaker could establish that a space to the left represents one character, and another space to the right represents a different character. The speaker can then refer to one of the spaces when wanting to refer to that character.

We have established that speakers produce gestures for self-oriented purposes, so how do we know that speakers also produce gestures to communicate? It is difficult to argue that gestures produced in isolation are produced for communicative purposes (e.g. Chu & Kita, 2008, 2011). However, gestures produced in conversational settings could have been produced for a communicative or self-oriented function. Gestures can be produced that serve more than one function. It is possible that a gesture intended to help the speaker think about an object in 3D space could help an onlooker think about the object. Furthermore, it is possible that if a speaker produces more gestures to convey information to an addressee, those gestures could also help the speaker prepare the information being discussed for speech. For example, when gestures can't be used to communicate, speakers produce fewer gestures, leading the speakers to become less fluent (Alibali et al., 2001). Studies can distinguish the independent function of communicative and self-oriented gestures by creating situations where participants should produce fewer gestures if gestures are only produced for self-oriented purposes, but participants produce more gestures (because the gestures are being produced for communicative purposes e.g. Holler et al., 2011). In a task where speakers needed to direct a partner to place cards into correct squares over six trials, the participants could build up a shared knowledge about the task. Familiarity with the task and this shared knowledge makes the task easier over time, meaning fewer gestures should be produced over time due to the reduced cognitive demands. However, speakers produced more gestures over trials. The findings suggest that gestures were used to coordinate referential communication between the speaker and addressee, fulfilling a communication function rather than just a self-oriented one. Research into the communicative function of gesture also comes in two categories: studies measuring the production of speaker's gesture and studies measuring the addressee's understanding of the information the speaker has encoded in gesture.

2.4.1 Producing communicative gestures to convey information

Speakers produce more gestures when the addressee can see the speaker's gestures. Speakers produce more gestures overall when speaking to a visible addressee rather than an non-visible addressee (Emmorey & Casey, 2001; Hostetter & Potthoff, 2012; Krauss et al., 1995; Mol et al., 2009a, 2009b). Furthermore, speakers produce more representational or illustrating gestures (Alibali et al., 2001; Cohen, 1977; Cohen & Harrison, 1973) and more interactive gestures (Bavelas et al., 1992, 2008) when speaking to a visible addressee rather than an non-visible. The difference in gesture rate has been attributed to the communicative function of gestures. When the addressee cannot see the speaker's gestures, then speakers are not motivated to communicate using gesture. Furthermore, speakers produce larger representational gestures when the addressee is visible (Bavelas et al., 2008). Speakers also placed gestures higher up in space when the addressee was visible (Holler et al., 2011). Speakers may produce gestures they wish to help communicate higher up in front of them and make them bigger. The larger, higher up gestures may be clearer and help the speaker convey information. Taken together, the findings suggest that when the addressee can see the speaker's gestures, then speakers are motivated to communicate using gesture.

The effect of visiblity on gesture can be mediated by individual differences and the communicative task. Extroverted speakers produce more gestures during communication (Hostetter & Potthoff, 2012). Furthermore, speakers who were introverted showed a larger effect of visibility than extroverted speakers when describing words (Hostetter & Potthoff, 2012). That is, introverts produced a much higher rate of gestures when the adressee was visible than when the addressee was not visible. On the other hand, extraverts producted marginally higher rate of gestures when the adressee was visible than when the addressee was not visible. The finding was explained by extraverts being more energetic in social situations and devoting the energy to gesturing, especially if gesturing frequently is a habit

(Hostetter & Potthoff, 2012). However, speakers who were extraverted showed a larger, reversed effect of visibility than introverted speakers when describing objects (O'Carroll et al., 2015). That is, introverts produced a a similar rate of gestures when the adressee was visible and when the addressee was not visible, but extraverts producted a lower rate of gestures when the adressee was visible than when the addressee was not visible. The finding were explained by the distinct needs of the different tasks. In the study by Hostetter & Potthoff (2012), speakers may have used gesture to help convey the appropriate word to the addressee, therefore speakers produced more gestures when the addressee was visible. In the study by O'Carroll et al. (2015) speakers may not have used gesture to communicate information to the addressee. Instead, speakers may have primarily used gesture to help the speaker find the right words to describe the objects. In both studies, the tasks were more difficult when the addressee was not visible. Extroverted speakers will not have been able to use feedback from the addressee to help them, and so gestured more in the difficult situation of the addressee not being visible. Taken together, the findings suggest that speakers use gesture differently in different situations, depending on the information being conveyed and the individual differences between speakers.

The effect of visibility on gesture may be attentuated by the type of information being conveyed by the speaker (Hostetter, 2014). According to the gesture as simulated action framework (GSA), gestures are the manifestations of action components in spatio-motoric simulations (Hostetter & Alibali, 2008, 2019). According to the GSA, speakers produce more gestures when the addressee is visible because speakers lower the threshold for the required motoric simulation activity (Hostetter, 2014). Speakers may do so when gestures would be communicatively useful (i.e. visible to the addressee). Speakers were asked to describe nouns to the addressee. The nouns had been rated on how manipulable the objects were. An item was rated manipulable by participants if they could imagine touching, holding, and using the item in a specific way. For example, hammer was rated as the most manipulable, and sunshine was rated as the last manipulable. The effect of visibiltiy was found to be the strongest for words with a low manipulability rating, and weakest for words with a high manipulability rating. When describing words which evoked high levels of motor activity (i.e. hammer), speakers produced gestures at a marginally higher rate when the addresse was visible. When describing words which evoked low levels of motor activity (i.e. sunshine), speakers produced gestures at a much higher rate when the addresse was visible than when the addressee was not visible. The findings suggest that the motor characteristics of the high manipulability words led to such high levels of activation that the gesture was produced regardless of whether the speaker raised the gesture threshold or not. The low manipulability words led to lower levels of motor activation, which passed the gesture threshold less often when the threshold was raised by the speaker than when the threshold was not raised. As the threshold is higher when the addressee is not visible, speakers produce fewer gestures.

Speakers produce more gestures in response to addressee feedback. Speakers produce more co-speech gestures when talking to responsive addressees than non-responsive addressees (Beattie & Aboudan, 1994). When addressees provide feedback to the speaker, then speakers produce more gestures in response to the addressee to confirm the message is understood correctly (Bavelas et al., 2011). Speakers also produce fewer gestures when communicating with a less attentive addressee (Jacobs & Garnham, 2007). Speakers and addressees can provide information back and forth between them to ensure the communication is effective (Schober & Clark, 1989). The speaker communicates a piece of information, the addressee can signal their understanding, and the speaker can confirm if the understanding is correct. Addressees can provide verbal responses such as “mm-hmm” and “yeah” as part of this back channel of communication between the speaker and addressee (Yngve, 1970). Addressees can also provide non-verbal responses (Bavelas et al., 2011). Speakers can send messages through back-channel responses using gaze (Bavelas et al., 2002). Furthermore, addressees can use facial expressions to respond to the speaker. These include smiles (Brunner, 1979) and mimicking other expressions (Bavelas, 2007; Bavelas et al., 2000). Speakers can then use gesture to respond to the addressee’s feedback during conversation to avoid having to interrupt themselves by responding vocally (e.g. Bavelas & Gerwing, 2011).

Speakers use different gestures to manage the conversation with the addressee. Speakers can indicate that they have finished speaking or wish to continue speaking using turn taking signals (Duncan, 1972). Speakers may return their hands to a resting position to signal they have finished speaking. Speakers may hold up their hand, palm facing the addressee, to signal they are supressing an attempt by the addressee to take a turn at speaking. Speakers can also produce interactive gestures to manage the interaction with the addressee (Bavelas et al., 1992). Interactive gestures can cite the previous contributions of the addressee, seeking a response from the addressee such as asking for help, delivering new information, conveying a lack of understanding, and signifying whose turn it is to speak (Bavelas et al., 1995). Speakers can also use gesture help convey information and manage the conversation by foreshadowing what the speaker will say or do immediately after

(Streeck, 2009). This way, the addressee can anticipate how the interaction will proceed, enabling quick and efficient co-ordination.

Speakers produce gestures differently depending on the social context, such as in high stakes situations. Speakers produce more gestures when communicating the information successfully is important (Kelly et al., 2011). Participants produced three times as many gestures when describing items in a survival kit for an informational video to students preparing for a rugged camping trip in the mountains than when describing the items to college students in an orientation activity. Mothers also produce a higher rate of gestures when conveying dangerous information than non-dangerous information to their child (Hilliard et al., 2015). The findings around gesturing in high stakes situations suggests that speakers are sensitive to the anticipated consequences of their communication and produce more gestures to communicate more effectively.

2.4.2 The effect of gestures on addressee comprehension

Evidence for the communicative function of gesture also comes from the effect of gesture on addressee comprehension. The consensus of the literature is that addressees better understand the speaker's message when the speaker is allowed to gesture than when the speaker is prohibited from gesturing. In two comprehensive meta-analyses investigating the communicative effect of gestures, speakers communicated more effectively when producing gestures than when gestures are prohibited (Dargue et al., 2019; Hostetter, 2011). The findings covered a wide array of situations where gesturing was effective. Addressees were better able to reproduce line drawings from descriptions given by speakers who gestured than from descriptions given by speakers who were prohibited from gesturing (Graham & Heywood, 1975). Children were better able to learn from teachers who produced gestures than teachers who did not produce gestures (Valenzeno et al., 2003). Addressees took in more information from adverts where a speaker was producing iconic gestures than adverts using only audio, or tv adverts where information was conveying through pictures (Beattie & Shovelton, 2005). Overall, gestures provide a moderate benefit to the addressee's understanding (Dargue et al., 2019; Hostetter, 2011).

Gestures that convey information which was not present in speech were more effective at conveying information to the addressee than gestures which conveyed information which was redundant with speech (Hostetter, 2011). Gestures which convey information not present in speech can be referred to as non-redundant, supplementary, or mismatching gestures (Church & Goldin-Meadow, 1986; Emmorey & Casey, 2001).

Addressees are better able to pick up information from the speaker when the speaker produced non-redundant gestures than when the speaker doesn't produce gestures at all (Church et al., 2007). Furthermore, addresses are able to answer questions directly related to the information only expressed in gesture (Beattie & Shovelton, 1999; McNeill et al., 1994). The meta-analysis compared the effect of non-redundant and redundant gestures on addressee comprehension (Hostetter, 2011). Studies in which speakers produced non-redundant gestures found effects of gesture on addressees' comprehension than studies in which speakers produced redundant gestures (Hostetter, 2011). Taken together, the findings suggest that addressees can take up information expresses solely in gesture, and that receiving complementary information in speech and gesture provides an overall benefit to comprehension.

Gestures are more effective at conveying information when depicting motoric information such as actions or spatial dynamic information such as the movement of objects than when depicting abstract metaphorical information (Hostetter, 2011). Gestures are particularly suited to conveying spatial information rather than abstract information because gestures can efficiently convey action, shape, relative location of objects and the movement of objects (Hostetter, 2011). In the meta-analysis, addressees benefitted from seeing gestures conveying spatial and motoric gestures, but not from gestures conveying abstract information (Hostetter, 2011). Furthermore, addressees benefitted from seeing gestures conveying motoric information than from seeing gestures conveying abstract information. However, the findings can be explained in two different ways. Firstly, iconic gestures may convey more information than metaphorical gestures. Addressees may be able to understand the spatial and motoric information encoded in iconic gestures because gestures are better at conveying spatio-motoric information. Addressees may be less able to take up the information from metaphorical gestures as gesture is less suited to conveying abstract information. Secondly, speakers produce more gestures when discussing spatio-motoric topics than when discussing abstract topics (Alibali, 2005; Beattie & Shovelton, 1999; Rauscher et al., 1996). The quantity of gestures produced when conveying spatio-motoric information could explain the benefit to addressees' comprehension. Further research is needed to distinguish between a gesture quality and gesture quantity explanation. However, both explanations suggest that gestures convey information to the addressee.

Gestures have been found to activate addressees' perceptual-motor networks, semantic networks, and social emotive processes in neurological studies (Yang et al., 2015). Gesture perception activated brain regions associated with the action-observation network.

The network may be used to understand of the actions and goals of others (Nummenmaa et al., 2014). In particular, the left ventral premotor cortex is part of the mirror neuron system, and is thought to be associated with connecting observed actions with our own action and motor representations (Rizzolatti, 2005). This connection may help with the interpretation of gestures. The perception of both co-speech and silent gestures activated brain regions associated with retrieval of lexical-semantic information during word processing (Yang et al., 2015). The left posterior middle temporal gyrus is activated when processing both iconic gestures (Willems et al., 2009) and metaphorical gestures (Kircher et al., 2009). The same neural processes responsible for the retrieval of lexical-semantic information from words may be responsible for the retrieval of semantic information from gestures.

2.4.3 Do speech and gesture compensate for one another?

Researchers have debated whether speakers use gestures to compensate for difficulties with verbal communication. According to the trade-off hypothesis (de Ruiter et al., 2012), speakers will rely more on gesturing when speech becomes difficult (de Ruiter, 2006), and speakers will rely more on speech when gesturing becomes difficult (e.g. Bangerter, 2004; Melinger & Levelt, 2004). An example of the speech-gesture trade-off is that speakers are prohibited from gesturing while discussing line drawings, the vividness of the spatial imagery in speech increased in comparison to participants who could gesture (Graham & Heywood, 1975). This finding suggested that speakers who could gesture conveyed some spatial imagery in gesture, when gesture was best suited to convey the information. When speakers were prohibited from gesturing, speakers would have to rely solely on speech to convey the information, even if gesture would have been better suited for communicating it. However, the trade-off hypothesis conflicts with the idea that gesture and speech are produced hand in hand. Speaker's gestures may parallel difficulties in speech rather than compensating for them (So et al., 2009). Speakers who referred to referents in gesture also frequently referred to the referents in speech. The finding contradicts the trade-off hypothesis as when the referent was identified using gesture, there would be no need to refer to the referent in speech. Furthermore, when the referent wasn't referred to using gesture, speakers would have needed to refer to the referent in speech to inform the addressee. The findings suggest that speakers produced gestures for cognitive self-oriented reasons (as we have seen earlier in the literature review) when talking to addressees. The gestures that speakers produce shape the speech, allowing speakers to be more fluent (Rauscher et al., 1996) and altering the way information is packaged for speaking (Kita et al., 2017). The trade-off hypothesis and the hand in hand hypothesis were directly compared experimentally (de Ruiter et al., 2012). The findings overall did not support the trade-off hypothesis, as the manipulations that made speech difficult affected speech but not gesture production. Gesture was not used to compensate for difficulty in speech. Furthermore, speakers who used more locative descriptions in speech also produced more pointing gestures than speakers who used few locative descriptions. This finding supported the hand in hand hypothesis rather than the trade of hypothesis. Taken together, the findings suggest that gesture and speech are produced hand in hand, though sometimes information that would normally be expressed only in gesture (such as some spatial information) is expressed in speech when gesture is suppressed (e.g. Graham & Heywood, 1975).

Taken all together, the findings suggest that speakers can produce gestures for communicative purposes. The literature suggests that speakers produce more gestures in situations where gesture is able to communicate information effectively to the addressee, such as when gestures can be seen (Alibali et al., 2001), or when gestures can convey visuo-spatial information (Alibali, 2005). Furthermore, the literature suggests that gestures are useful in communicating information to the addressee.

2.5 Audience design

Speakers can design their utterances to effectively convey a message to an addressee. Audience design refers to the ways in which speakers tailor their communication to benefit their addressee (Clark & Murphy, 1982). Speakers can exploit the common ground, or mutual knowledge, they share with their addressee in designing their utterances (Clark & Marshall, 1981; Krauss & Fussell, 1991). For example, speakers describe objects to clearly distinguish a target object from other similar objects if they know that the addressee can see these objects as well (Horton & Keysar, 1996). Speakers can also take feedback from the addressee into account to design future utterances for the addressee (Krauss & Fussell, 1991; Kuhlen & Brennan, 2010). Furthermore, speakers can shape their communication based on their knowledge about the addressee (Krauss & Fussell, 1991) such as whether the addressee is a child. For example, speakers use shorter sentences and simpler words when speaking to children, compared to other adults (Snow, 1972), and produce larger and more enthusiastic hand gestures (Brand et al., 2002).

2.5.1 The perspective taking mechanism

Audience design behaviours can be explained by a perspective taking mechanism. Speakers can form hypotheses about the addressee such as what the addressee knows and believes (Grice, 1975; Levelt, 1989). Speakers can use these hypotheses to design their communicative behaviours for the addressee. To take the perspective of the addressee, speakers build and maintain a model of the addressee and accesses this model when necessary (Clark & Marshall, 1981; Clark & Murphy, 1982). Speakers constantly update this model over time as new information about the addressee becomes available (Krauss & Fussell, 2004). This perspective taking mechanism is used to explain audience design behaviours, in which speakers combine beliefs about what information the addressee knows (previously stored in the model) and feedback from the addressee (used to update the model) to adjust communicative behaviours to better communicate with the addressee (Krauss & Fussell, 1991).

Speakers take the perspective of the audience at some point in utterance production. One view is that speakers only take the perspective of the addressee after planning their initial utterance and adjust afterwards (Horton & Keysar, 1996; Pickering & Garrod, 2004; Rossnagel, 2000). Another view is that speakers take the perspective of the addressee into account from the earliest moments of language production (e.g. Winner et al., 2019). Evidence suggests that the addressee's perspective can be incorporated early on (see Brennan & Hanna, 2009, for a review).

Some research has investigated the perspective taking mechanism by looking at how the speaker incorporates information into their model of the addressee. Speakers may build and update a model of the addressee by using heuristics such as the co-presence heuristic, where speakers assume addressees know something if the speaker and addressee were both present when that information was presented (Craycraft & Brown-Schmidt, 2018). That is, the co-presence heuristic can specify how speakers determine what information is in the common ground. Another heuristic is an egocentric anchoring and adjustment heuristic (Deliens et al., 2017; Epley et al., 2004). The speaker starts with their own egocentric perspective and adjusts from this initial anchor by making incremental changes. After each adjustment, the speaker evaluates whether this new perspective plausibly estimates the addressee's perspective. Most of the research into audience design has focussed on the perspective taking approach to audience design, but there has been little research into alternative mechanisms that do not involve a model of the addressee.

2.5.2 The cue-based heuristic mechanism

Audience design behaviours have also been explained by a cue-based heuristic mechanism. It may not be feasible for speakers to constantly take the addressee's perspective and utilise a model of the addressee. Maintaining a model of the addressee's knowledge is cognitively demanding (Rossnagel, 2000) and time consuming (Epley et al., 2004; Horton & Keysar, 1996). It may be too costly for speakers to constantly consult a model of the addressee's knowledge to produce audience design behaviours. An alternative cue-based heuristic mechanism could explain some audience design behaviours. When the speakers' behaviour seems to be tailored for their addressee, it does not necessarily mean that speakers designed those behaviours for the addressee (Barr & Keysar, 2006). Speakers can instead take advantage of cues available in their interaction with the addressee (Shintel & Keysar, 2009). Such cues can trigger speakers to change their behaviour in pre-determined ways. Speakers do not have to use these cues to first update a model of the addressee (as in the co-presence cue, discussed above) and then adjust communicative behaviour based on the model. Cues that directly trigger audience design behaviours might include addressee feedback, visual behaviours such as eye gaze and gesture, and the addressee's status as a child or foreign speaker.

There is little empirical evidence to unequivocally support the cue-based heuristic as a mechanism of audience design. Most support for the cue-based heuristic mechanism relies on the argument that heuristics are a cognitively economical alternative to the perspective taking mechanism (Galati & Brennan, 2010, 2014; Shintel & Keysar, 2009). However, the cognitive economy argument relies solely on parsimony, and not on empirical evidence (Osterhout & Swinney, 1989). To our knowledge no study has a design in which the perspective taking, and cue-based heuristic mechanisms make different predictions. Some research has investigated the two mechanisms, but the design has not allowed for distinct predictions for the perspective taking and cue-based heuristic mechanisms. For example, Blokpoel et al. (2012) investigated how speakers reacted to mistakes by an addressee during a communication game. Speakers would signal what moves the addressees should make on a board while the speaker is moving their own piece. Speakers would respond to addressees' mistakes by clarifying the signals in the next trial. For instance, speakers paused for a longer duration at the addressee's target location after addressees made a location error in the task (the addressee moved a piece to the wrong place on a board, but the piece was facing the right way). Speakers did not pause at the target location after addresses made an orientation error (the addressee moved a piece to the right place on a board, but the piece was facing the wrong way), or after addresses made a combined error (both location and orientation were wrong). The authors concluded that speakers take the perspective of the addressee because heuristics are too simple to explain the finding that speakers paused after addressees made location errors but not after addressees made location and orientation errors together. The authors argue that if heuristics need to be complicated enough to follow complex multiple part rules, the findings can more parsimoniously be explained by a perspective taking mechanism. However, the argument that cue-based heuristics can only be “one-bit” or one-rule (Brennan & Hanna, 2009), once again relies on arguments of parsimony. There is no reason to suggest that heuristics cannot be more than “one-bit”, and researchers who have attributed their findings to “one-bit” models have stated that further research is needed to establish whether more complex heuristic rules are needed to model audience design in human communication (e.g. Galati & Brennan, 2010).

2.5.3 Testing the mechanism behind audience design behaviours

We can establish an audience design effect as being explained by a perspective taking mechanism when the process of designing the behaviour is cognitively costly and time consuming. A key argument for the cue-based heuristic mechanism is that taking the perspective of the addressee is time consuming and cognitively costly. Therefore, if an audience design behaviour is observed when the speaker has the time and cognitive resources to take the perspective of the addressee, and not observed when the speaker is under time pressure or cognitive load, then the audience design behaviour is most likely the result of perspective taking. For example, when speakers were asked to give instructions to an addressee on how to build a machine model, speakers tailor the instructions to the addressee based on the addressee’s expertise (as either an adult or child) (Rossnagel, 2000). When the speakers were placed under high cognitive load (recalling the instructions from memory or remember a list of numbers), they no longer adapted their speech to consider the expertise of the addressee. The speakers used high rates of technical terms and non-specified names for parts, on a comparable level to when describing the building process to adults. The finding cannot be explained by a cue-based heuristic explanation as automatically responding to cues from the addressee would not be impacted by high cognitive load. Furthermore, establishing the expertise of an addressee at building mechanical models most likely requires making inferences about the knowledge of the addressee. The finding can instead be explained by a perspective taking mechanism. The speakers would have to develop a model of the addressee with information about the addressee’s likely knowledge of model making, before then consulting the model during utterance production to plan speech to communicate information at an appropriate level for the addressee. Audience design behaviours can also be attributed to perspective taking mechanisms when the behaviours disappear while speakers are under time pressures. For example, speakers make more egocentric judgements about potentially sarcastic statements when hurried during the task (Epley et al., 2004), and speakers are less likely to consider which information is in common ground while under time pressures (Horton & Keysar, 1996). Similarly to being under high cognitive load, speakers cannot take the perspective of the addressee under time pressures. It is therefore plausible that we can test if an audience design behaviour results from perspective taking by testing if the behaviour disappears under high cognitive load or restrictive time pressures.

We can establish an audience design effect as a result of a cue-based heuristic mechanism when the heuristic rules behind the behaviour can result in errors. Heuristics in decision making have been associated with the idea of systematic bias or errors from their conception (Tversky & Kahneman, 1974). For example, the availability heuristic leads people to judge the relative frequency of an occurrence by how easily past occurrences are to retrieve from memory (Schwarz et al., 1991). We are therefore biased into believing that a certain event is more likely than it really is if the event is more memorable than the alternatives. For example, participants often incorrectly state there are more words beginning with the letter ‘r’ in the English language than words with ‘r’ as the third letter (Tversky & Kahneman, 1973). The systematic error is the result of words beginning with ‘r’ being more easily recalled that words with ‘r’ as the third letter. The findings suggest that more words beginning with ‘r’ were available to participants upon recall, leading participants to presume those words were more common.

Systematic biases and errors should exist within audience design cue-based heuristic mechanisms. We can establish a behaviour is triggered by cues from the addressee by presenting the cues to the speaker to trigger audience design behaviours. If speakers produce the audience design behaviour in the presence of the cue regardless of whether the behaviour can be effective or not, then the behaviour is plausibly being produced in response to the cue rather than by the speaker making inferences about whether the audience design behaviour would benefit communication.

2.5.4 Gestural audience design behaviours

Speakers can also design their gestures to convey a message more effectively to the addressee, rather than only designing their speech. Speakers will orient their gestures to move within a space they share with their addressee when speaking to a single addressee, or within a central space when speaking to two addressees (Özyürek, 2002). Speakers adapt their communication by presenting some spatial information in gesture rather than speech, suggesting speakers will present information in gesture when gesture is best suited to communicate that information to the addressee (Melinger & Levelt, 2004). Speakers take their addressee's viewpoint into account when producing pointing gestures (Winner et al., 2019). Speakers attenuate their gestures when retelling the story to an addressee they had already told the same story (Galati & Brennan, 2014; Jacobs & Garnham, 2007). Speakers also make communicative adjustments in response to addressee behaviours. Speakers' gestures become more precise, larger, or more visually prominent after feedback (Holler & Wilkin, 2011). Speakers produce more gestures when an addressee could see the speaker's gestures (Alibali et al., 2001). Furthermore, speakers are more likely to express size information in gesture when the size information is new to the addressee than when the size information is already known to the addressee (Holler & Stevens, 2007). Taken together, speakers adjust the production of gestures to increase the effectiveness of communication to the addressee.

2.5.5 A mechanistic framework for audience design behaviours

Audience design behaviours can be explained by the mechanistic framework for audience designs (Ferreira, 2019). The mechanistic framework can be divided into three key processes: feedforward language production, recurrent forward model processing, and a learning process. The feedforward language production process is made up of the message encoding, grammatical encoding, and phonetic coding stages. These processes take the speaker from their communicative intentions to having a formulated pre-articulatory utterance. Speakers can design the utterance for the addressee by exerting executive control over the message encoding process to adapt the message, and phonetic encoding process to adapt the sounds. An example of this feedforward audience design is the Lombard effect (Lombard, 1911). Speakers can exert executive control over the phonetic encoding process to increase the volume of the utterance while a loud noise is present. Feedforward audience design can only consider information known to the speaker at the beginning of speech production during the message encoding stage. The early stages of the feedforward language production process pass on the communicative intention and the message to the recurrent forward model process. The forward model determines the communicatively relevant linguistic features and simulated how the addressee might comprehend the message. The output of the simulation is then compared against the communicative intention of the speaker. If the predicted communicative effect and the speaker's intention match, then no further changes need to be made. If the predicted communicative effect and the speaker's intention do not match, the speaker can exert executive control over the feedforward language production process to compensate. Finally, the learning component explains how the feedforward language production process obtains audience design strategies from the recurrent forward model process. Speakers can observe the actual communicative effect of a produced utterance and compare it to the predicted communicative effect from the forward model process. Rules can be build based on the discrepancies between the actual and predicted communicative effects. These rules can be incorporated into the feedforward production process enabling speakers to avoid the same mistakes in the future by exerting executive control over the process.

The perspective taking and cue-based heuristic views on audience design can both be incorporated into the mechanistic framework for audience design. The perspective taking mechanism can explain how speakers simulate the addressee's comprehension during the recurrent forward model process. Once speakers have generated the relevant linguistic features during the forward modelling stage, those features can be passed through a simulation where speakers take into account their model of the addressee (Clark & Marshall, 1981; Clark & Murphy, 1982). By consulting the model of the addressee, speakers can infer whether the predicted communicative outcome will match the communicative intention. The cue-based heuristic mechanism can explain how speakers use rules previously learned to alter utterances during the feedforward production process. Plausibly, the audience design behaviours learned during the recurrent forward model process could be cue-based, with speakers learning rules/heuristics to solve repeated discrepancies between the actual and predicted communicative effects. Once a rule has been learned, speakers can produce the audience design behaviour during the feedforward language production process.

While the mechanistic framework for audience design is created to explain verbal audience design behaviours, it is plausible that the same system can account for non-verbal audience design behaviours. As discussed previously, gestures are thought to be produced by processes responsible for speech production (Goldin-Meadow, 1999; McNeill, 1992), in conjunction with the action generation process (Chu & Kita, 2016; Hostetter & Alibali, 2008, 2019; Kita, 2000). If gestures are to be produced as part of a feedforward language production process, then speakers should be able to exert executive control over specific gesture generation processes. The message encoding process would be the same process for speech and gesture. At this stage, the process would determine what information form the message will be expressed in gesture. Rather than a grammatical encoding process, gesture could undergo an equivalent process where the form of the gesture is determined. For instance, the process could determine the speaker will express the idea of 'apple' by miming biting into an apple. Rather than a phonetic encoding process, gesture could be subject to a process determining the size and location of the gesture. For instance, speakers could exert executive control over this process by making a gesture larger if the information being conveyed by the gesture is important. The recurrent forward model process could take into account gesture when determining communicatively relevant 'linguistic' features, integrating speech and gesture before comparing the speaker's communicative intention with the predicted communicative effect.

Taken together, the findings in the audience design literature suggest that speakers take the perspective of their addressees to infer which behavioural changes. No prior research has provided empirical evidence for a cue-based heuristic explanation. Previous findings have been reinterpreted as supporting a heuristic mechanism (Shintel & Keysar, 2009), but the findings can be explained by both the perspective taking view and the cue-based heuristic view. I have laid out how research going forward can differentiate perspective taking and heuristic triggered audience design behaviours. The views can be investigated by measuring the impact of cognitive load and time restraints, and by exploiting potential systematic errors and biases in heuristics. Furthermore, the two mechanisms do not have to be mutually exclusive. The mechanistic framework for audience design (Ferreira, 2019) can how both the perspective taking and cue-based heuristic mechanisms can be responsible for audience design behaviours. In addition, the findings in the literature suggests that audience design behaviours are not limited to verbal utterances, but that speakers adjust non-verbal communication such as gestures as well. There is nothing to suggest that the mechanism behind gestural audience design effects differs from the mechanism responsible for verbal audience design behaviours. Based on these findings in the literature, I can examine the mechanisms responsible for audience design effects in gesture production.

2.5.6 *Foregrounded Gestures*

Speakers can place gestures in the foreground of the interaction to help communicate information using gesture. Speakers can design their gestures to attract the addressee's attention and clearly convey important information (Cooperrider, 2018). Cooperrider (2018) defines foreground gestures as being in the forefront of both the speaker's and the addressee's awareness, as well as being in the foreground of the interaction between the speaker and addressee. When speakers produce foreground gestures, they can provide a clear indicator that the speaker produced the gesture to communicate information to the addressee (Streeck, 1993).

Speakers can put more effort into their gestures to signal gestures use in communication. Speakers may make their gestures larger to signal the gestures use for communication (as suggested by Cooperrider, 2018) and to make the gesture more clear to the addressee (as suggested by Bavelas et al., 2008). Speakers can put more effort into their gestures by making the gestures larger (and more visually prominent) or by making gestures more precise. Gesture size has been examined as an outcome variable in gesture research. For instance, speakers produce larger representational gestures when the addressee is visible (Bavelas et al., 2008). Speakers also placed gestures higher up in space when the addressee was visible (Holler et al., 2011). Speakers produced larger gestures when cooperating with the addressee than when competing with the addressee. Speakers produce larger and more visually prominent in response to feedback from the addressee (Holler & Wilkin, 2011). The findings suggest that speakers can make their gestures larger to help communicate information to the addressee via gesture. Speakers can also make their gestures more precise to signal the gestures use for communication (as suggested by Cooperrider, 2018) and so that the gesture is clearer in conveying information to the addressee (as suggested by Gerwing & Bavelas, 2004). Gestures are found to be more precise when speakers are conveying new information to an addressee than when conveying information already in common ground to an addressee (Gerwing & Bavelas, 2004). Furthermore, speakers produce more precise gestures in response to feedback from the addressee (Holler & Wilkin, 2011). The finding suggests that speakers use more precise gestures whenever the speaker wishes to communicate information to the addressee via gesture. However, there are no findings to suggest that producing larger or more precise gestures benefits the addressee's comprehension. Furthermore, there are no findings that larger or more precise gestures capture the attention of the addressee.

Speakers can fixate their gaze on their own gestures to signal the gestures use in communication. Speakers are thought to coordinate their gaze with gestures when producing gestures intended for communication (Enfield, 2009; C. Goodwin, 1986; Streeck, 1993). When speakers are fixating on their own gestures, speakers must be fully aware of their own gestures they are producing, which signals to the addressee that the gestures are intended to convey information. Speakers both fixate their gaze on gestures and indicate their gestures using speech when speaking to children (Slonimska et al., 2015). Despite numberous observations (e.g. Streeck, 1993), there is no unequivical evidence that speakers gaze at their gestures more when trying to communicate information using gesture. For example, there is no established effect of visibility on speakers gaze fixation behaviour. However, findings suggest that speakers fixating their gaze on their own gestures affects addressee behaviour and comprehension. Addressee's attention is drawn to the speaker's gestures when speakers fixate on their own gestures (Gullberg & Kita, 2009). Furthermore, speakers fixating on their gestures did benefit addressee's comprehension of directional information (Gullberg & Kita, 2009). However, only an effect of gaze fixation on addressee's recall of direction information was tested and the effect was not reproduced when speech was controlled for using video editing to create artificial gaze fixations on gestures. Furthermore, gesture duration may have been confounded with gaze fixation as gestures were longer when the speaker fixated on the gestures that when the speaker did not fixate on the gestures. While the findings suggest the speakers fixating gaze on gestures benefits addressee comprehension, further work can be carried out to expand the findings from only investigating directional information and the potential confounding of gaze fixation, gesture duration, and speech production.

Speakers can use refer to their gesture in speech using demonstratives (Guérin, 2015). Demonstratives can include phrases such as 'like this' or 'like that'. Speakers use more demonstratives when pointing to closer objects (in this case faces) than when pointing to further away objects (Bangerter, 2004). When the objects are closer, the gesture is easier to interpret so speakers will draw attention to the gesture. When the objects are further away, the gesture is difficult to interpret, speakers do not draw attention to the pointing gesture and rely instead on speech to clarify. The finding suggests that speakers draw attention to gestures using speech more when trying to communicate information using gesture. Furthermore, speakers use more demonstratives alongside gesture when the gestures are visible to the addressee than when gestures are not visible to the addresee (Bavelas et al., 2008; de Ruiter et al., 2012; Emmorey & Casey, 2001). Speakers are also especially likely to use demonstratives in association with gaze fixation behaviours when talking to children (Slonimska et al., 2015). Taken together, the findings suggest that speakers use demonstratives to highlight gestures whenever the speaker wishes to communicate information to the addressee via gesture. This is especilly the case when talking to children, as the addressee can direct the child's attention to gesture. However, there are no findings to suggest that producing demonstratives alongside gestures benefits the addressee's comprehension, nor that producing demonstratives alongside gestures draws the attention of the addressee to the speaker's gestures.

Note for the empirical chapters

Chapters 3, 4, and 5 are the empirical chapters for this thesis. These chapters are based on self-contained manuscripts being prepared for publication. Therefore, each chapter has it’s own abstract, introduction, and discussion/general discussion sections. There will be some overlap between these sections in the experical chapters and the thesis introduction and discussion. Any mentions of supplimentary materials in the experimental chapters can be found in the appendices.

Chapter 3

The Addressee's Face as a Trigger for Gesture Production: Evidence for Cue-Based Heuristics in Audience Design

Abstract

Speakers can design their verbal utterances and gestures to better convey information to an addressee; such adjustment of communication is called audience design. Audience design can be achieved by speakers taking the perspective of the addressee, considering what the addressee knows, believes, and can see. However, perspective taking can be a cognitively intensive and time-consuming process. Speakers may also use simple cues from the addressee to trigger some audience design behaviours. The current study investigated whether speakers could utilise cue-based heuristics when using gesture to communicate. We hypothesised that the addressee's face can act as a cue for the speaker to produce more gestures. Speakers indeed produced more gestures when the addressee's face could be seen than when the addressee's face could not be seen, but only when describing dynamic spatial stimuli (Experiment 2), not when discussing abstract stimuli (Experiment 1). It was not the case that speakers produced more gestures when the addressee could see the speaker's gestures than when the addressee could not (Experiment 3); that is, speakers did not take the addressee's perspective to determine that the addressee could not see the speaker's gestures. Thus, speakers use the addressee's face as a cue to trigger additional gestures, but only when discussing spatial stimuli for which gestures are likely to be useful for conveying a message. The current study provided unequivocal evidence for cue-based audience design in communication.

Keywords: recipient design, mental shortcut, Non-verbal communication, communicative adjustment, perspective taking

3.1 Introduction

Speakers often design their utterances to effectively and efficiently convey their message to their addressee. Audience design refers to the ways in which speakers tailor their communication to benefit their addressee (Clark & Murphy, 1982). Speakers can exploit the common ground, or mutual knowledge, they share with their addressee in designing their utterances (Clark & Marshall, 1981; Krauss & Fussell, 1991). For example, speakers describe objects to clearly distinguish a target object from other similar objects if they know that the addressee can see these objects too (Horton & Keysar, 1996). Speakers can also take feedback from the addressee into account to design future utterances for the addressee (Krauss & Fussell, 1991; Kuhlen & Brennan, 2010). Furthermore, speakers can shape their communication based on their knowledge about the addressee (Krauss & Fussell, 1991) such as whether the addressee is a child. For example, speakers use shorter sentences and simpler words when speaking to children, compared to other adults (Snow, 1972), and produce larger and more enthusiastic hand gestures (Brand et al., 2002). The current study investigated such audience design behaviours, that is, how adult speakers change their communication to benefit their addressee.

There are two types of audience design behaviours: reflexes and communicative adjustments. The Lombard effect (Lombard, 1911), which is the involuntary tendency of speakers to increase their vocal effort when speaking in loud noise, can provide examples of both types of audience design. The Lombard effect can exemplify a reflexive audience design behaviour because the effect is produced in response to an environmental cue that is not related to the addressee. Speakers told stories louder when there was noise despite being alone in the room (Pick et al., 1989), suggesting that the Lombard effect is in response to the noise rather than an addressee being unable to hear the speaker. This reflexive behaviour can still be classified as an audience design behaviour because it would benefit any addressees present. The Lombard effect can also exemplify communicative adjustment because speakers' modulation in response to background noise was greater when an addressee was present in the room, listening to them (Garnier et al., 2010). This indicates that the Lombard effect is more than just a reflex. This type of Lombard effect is a communicative adjustment because speakers change their behaviour in response to cues related to the addressee. The current study concerns the latter type of the audience design behaviours: communicative adjustments.

3.1.1. A perspective taking mechanism for communicative adjustments

Speakers can adjust communicative behaviour by taking the perspective of the addressee. Speakers can form hypotheses about the addressee such as what the addressee knows and believes (Grice, 1975; Levelt, 1989). Speakers can use these hypotheses to design their communicative behaviours for the addressee. To take the perspective of the addressee, speakers build and maintain a model of the addressee and accesses this model when necessary (Clark & Marshall, 1981; Clark & Murphy, 1982). Speakers constantly update this model over time as new information about the addressee becomes available (Krauss & Fussell, 2004). This perspective taking mechanism is used to explain audience design behaviours, in which speakers combine beliefs about what information the addressee knows (previously stored in the model) and feedback from the addressee (used to update the model) to adjust communicative behaviours to better communicate with the addressee (Krauss & Fussell, 1991).

Some research has investigated the perspective taking mechanism by looking at how speakers incorporate information into their model of the addressee. Speakers may build and update a model of the addressee by using heuristics such as the co-presence heuristic, where speakers assume addressees know something if the speaker and addressee were both present when that information was presented (Craycraft & Brown-Schmidt, 2018). That is, the co-presence heuristic can specify how speakers determine what information is in the common ground. Another heuristic is an egocentric anchoring and adjustment heuristic (Deliens et al., 2017; Epley et al., 2004). Speakers start with their own egocentric perspective and adjusts from this initial anchor by making incremental changes. After each adjustment, speakers evaluate whether this new perspective plausibly estimates the addressee's perspective. Most of the research into audience design has focussed on the perspective taking approach to audience design, but there has been little research into alternative mechanisms that do not involve a model of the addressee.

3.1.2. A cue-based heuristic mechanism for communicative adjustments

It may not be feasible for speakers to constantly take the addressee's perspective and utilise a model of the addressee. Maintaining a model of the addressee's knowledge is cognitively demanding (Rossnagel, 2000) and time consuming (Epley et al., 2004; Horton & Keysar, 1996). It may be too costly for speakers to constantly consult a model of the addressee's knowledge to produce audience design behaviours. Therefore, some audience design behaviours may not be the result of perspective taking, but instead are the result of a less cognitively demanding and time-consuming process.

An alternative cue-based heuristic mechanism could explain some audience design behaviours. When the speakers' behaviour seems to be tailored for their addressee, it does not necessarily mean that speakers designed those behaviours for the addressee (Barr & Keysar, 2006). Speakers can instead take advantage of cues available in their interaction with the addressee (Shintel & Keysar, 2009). Such cues can trigger speakers to change their behaviour in pre-determined ways. Speakers do not have to use these cues to first update a model of the addressee (as in the co-presence cue, discussed above) and then adjust communicative behaviour based on the model. Cues that directly trigger audience design behaviours might include addressee feedback, visual behaviours such as eye gaze and gesture, and the addressee's status as a child or foreign speaker.

Some audience design behaviours could potentially be driven by a cue-based heuristic mechanism. One example is story attenuation (Galati & Brennan, 2010). When speakers retell a story for a second or third time, speakers attenuate (i.e., simplify) the story they were retelling when they had already told the story to the same addressee. Speakers mentioned fewer events, used fewer words per event and provided less detail. According to a heuristic view, speakers check if they are telling this story to the addressee for the first time or have told them before. The cue from the identity of the addressee triggers speakers to attenuate the linguistic formulation of the story. While the cue-based heuristic mechanism can explain this effect, the perspective taking mechanism can also explain it. Speakers could take the addressee's perspective and infer that because the addressee has been told the story before, they do not need as much detail of the story.

There is little empirical evidence to unequivocally support the cue-based heuristic as a mechanism of audience design. To our knowledge no study has a design in which the perspective taking and cue-based heuristic mechanisms make different predictions. Some research has investigated the two mechanisms (e.g. Blokpoel et al., 2012), but the design has not allowed for distinct predictions for the perspective taking and cue-based heuristic mechanisms. Most support for the cue-based heuristic mechanism relies on the argument that heuristics are a cognitively economical alternative to the perspective taking mechanism (Galati & Brennan, 2010, 2014; Shintel & Keysar, 2009). However, the cognitive economy argument relies solely on parsimony, and not on empirical evidence (Osterhout & Swinney, 1989). The current study aims to provide direct empirical evidence to support cue-based heuristics as a mechanism of audience design.

3.1.3. The use of audience design in gesture

Audience design behaviours are not exclusive to verbal communication and are prevalent in non-verbal communication such as gestures. Speakers produce co-speech gestures when they talk, which are the hand and arm movements spontaneously made during speech (McNeill, 1992). Speakers can design co-speech gestures for addressees. Speakers will orient their gestures to move within a space they share with their addressee when speaking to a single addressee, or within a central space when speaking to two addressees (Özyürek, 2002). Speakers adapt their communication by presenting some spatial information in gesture rather than speech, suggesting speakers will present information in gesture when gesture is best suited to communicate that information to the addressee (Melinger & Levelt, 2004). Speakers take their addressee's viewpoint into account when producing pointing gestures (Winner et al., 2019). Speakers attenuate their gestures when retelling the story to an addressee they had already told the same story (Galati & Brennan, 2014; Jacobs & Garnham, 2007). Speakers also make communicative adjustments in response to addressee behaviours. Speakers' gestures become more precise, larger, or more visually prominent after feedback (Holler & Wilkin, 2011). Speakers produce more gestures when an addressee could see the speaker's gestures (Alibali et al., 2001). Taken together, speakers adjust the production of gestures to increase the effectiveness of communication to the addressee.

None of the findings into the gestural audience design behaviours unequivocally support a cue-based heuristic mechanism. Similarly to verbal communication studies, no non-verbal communication studies had a design that distinguishes the predictions of a perspective taking and cue-based heuristic mechanism. Cue-based heuristics can account for the findings of gestural audience design behaviours (e.g. Galati & Brennan, 2014), but as with verbal communication studies, authors rely on the argument of parsimony to motivate an explanation based on a heuristic mechanism. To overcome this limitation, the current study aims to provide unequivocal empirical support for the cue-based heuristic mechanism.

The current study investigated the mechanism underlying the effect of visibility on gesture production. Speakers produce more gestures when an addressee can see the speaker than when the addressee cannot (e.g. Alibali et al., 2001). This effect can be explained based on perspective taking and cue-based heuristics. According to a perspective taking view, the speaker takes the addressee's perspective to infer if the addressee can see the speaker or not and produce more gestures if the addressee can see the gestures.

According to the cue-based heuristic view, speakers produce more gestures in reaction to a cue related to the addressee. The addressee's face being visible could be such a cue. When speakers see the addressee's face, a heuristic process could be triggered and increase the rate of gesture production.

The addressee's face may be a cue for triggering gesture production because being able to see each other's faces is an important part of conversation. Addressees typically fixate constantly on the speaker's face during conversation (Gullberg & Holmqvist, 2006). Speakers are sensitive to the gaze of the addressee and use speech and gesture to draw the addressee's gaze towards the speaker (C. Goodwin, 1986; M. H. Goodwin & Goodwin, 1986). Addressees use facial expressions to provide feedback to the speaker, which can impact the speaker's communication (Bavelas et al., 2000; Bavelas & Gerwing, 2011).

3.1.4. The Current Study.

The current study addressed the lack of empirical evidence for the cue-based heuristic view of audience design. As previously stated, the literature has argued for a cue-based heuristic mechanism on the theoretical basis that heuristic mechanisms are simpler and more cognitively efficient (e.g. Shintel & Keysar, 2009). Prior studies investigating audience design do not have designs in which the perspective taking and cue-based heuristic mechanisms make different predictions. The current study will investigate the effect of visibility on gesture and make distinct predictions for the perspective taking and cue-based heuristic mechanisms

The current study investigated whether a heuristic or perspective taking mechanism is responsible for the effect of visibility on gesture. We hypothesise that a heuristic mechanism would utilise the addressee's face as a cue for speakers to use gesture for communication. More specifically, speakers should produce more gestures when there is no visual barrier between the speaker and the addressee because they can see the addressee's face, but not because the addressee can see the speaker's gestures.

We aim to distinguish between the perspective taking and cue-based heuristic mechanisms by separating facial visibility and gesture visibility in our experiments. Previous studies have manipulated gesture visibility using a screen to prevent the speaker and addressee from seeing one another (e.g. Alibali et al., 2001). In addition to the conventional manipulation of visibility, the current study used a novel condition in which a shoulder height screen was placed between the speaker and the addressee to prevent gestures being seen by the addressee while allowing the speaker and addressee to see each other's faces.

We compared speakers' gesture rate and gesture height in three conditions. In a first condition the speaker communicated with the addressee face to face with no screen, so the speaker and addressee could see each other without any restrictions. In a second condition the speaker communicated with the addressee with a shoulder height screen between them, so the addressee could not see the speaker's gestures, but the speaker and addressee could see each other's faces. In a third condition the speaker communicated with the addressee with a full height (above head) screen between them, the speaker and addressee could not each other at all.

Previous studies have used a variety of stimuli to investigate the effect of visibility on gesture. Some studies have used abstract stimuli (e.g. Bavelas et al., 1992; Rimé, 1982). Others have used spatial stimuli (e.g. Alibali et al., 2001; Mol et al., 2009): either static spatial stimuli such as line drawings (e.g. Bavelas et al., 2014) or dynamic spatial stimuli such as cartoons (Alibali et al., 2001). In Experiment 1 we will test our hypotheses when speakers discuss abstract stimuli and in Experiment 2, we will test our hypotheses when speakers describe dynamic spatial stimuli. We will conduct an exploratory analysis as part of Experiment 2 to see if the effect of visibility on gesture differs depending on the topic speakers are talking about.

We distinguished different types of gestures. Representational gestures depict an action, movement, object, shape, location, or metaphorically depict abstract concepts. Representational gestures are produced more often when the addressee can see the speaker's gestures (Alibali et al., 2001). Interactive gestures support the interaction with the addressee and directly reference the addressee. Interactive gestures are also produced more often when the addressee can see the speaker's gestures (Bavelas et al., 1992). We also measured the height of the speaker's representational gestures to check if speakers are producing their gestures higher in the shoulder height screen condition than in the no screen condition as this would indicate speakers are gesturing higher so the addressee can see the speaker's gestures.

According to the heuristics view, speakers should produce more gestures whenever the speaker can see the addressee's face. We predict that speakers should produce both representational and interactive gestures at a higher rate in the no screen and shoulder height screen conditions than in the full screen conditions. In addition, the heuristics view predicts that speakers will not move their gestures higher in the shoulder height screen condition than the no screen condition.

According to the perspective taking view, speakers should produce more gestures when the addressee can see the speaker's gestures. We can make two alternative predictions. The first prediction assumes that speakers do not raise their gestures above the shoulder height screen. In this case, speakers should produce both representational and interactive gestures at a higher rate in the no screen condition than the shoulder height and full screen conditions. The alternative prediction assumes that speakers raise their gestures above the shoulder height screen so the addressee can see the gestures. In this case, speakers should produce gestures at a higher rate in the no screen and shoulder height screen conditions than in the full screen conditions, and speakers should produce gestures higher in the shoulder height screen condition than in the no screen condition.

One alternative behaviour that speakers may display is using head movements when gestures cannot be seen. If the current study finds evidence for cue-based heuristics, it is possible that speakers did take the perspective of the addressee and rather than adjusting their hand gestures, speakers may have used head movements to convey information to the addressee. Speakers can use head movements to convey specific information (Stivers, 2008). Information can be conveyed not only through nods and shakes of the head, but also more complex representational movements (McClave, 2000). Speakers can use head movements to express the mental images of characters in a story, establish and refer to locations in space, listing alternatives, requesting feedback from the addressee and for taking on the role of a character when producing a quote (McClave, 2000). Head movements can also be used for managing turn taking in a conversation (Duncan, 1972). It is therefore plausible that speakers may compensate for not being able to easily produce visible hand movements by signalling equivalent information in head movements. In the current study, we measured speakers head movements whenever an effect of visibility was found to rule out speakers using head movements to compensate for the screen blocking the use of communicative gestures.

Regardless of which mechanism is responsible for the effect of visibility on gesture, to make sure that the addressee's behaviour is not responsible for the effect, we manipulated the addressee's responsive behaviour. Speakers produce more gestures when the addressee is responsive than unresponsive (Beattie & Aboudan, 1994). In the full screen condition, the addressee cannot see the speaker's gestures, but also the speaker cannot see the addressee's visual responses such as facial expressions to show understanding or confusion. Without the visual responses from the addressee, the speaker may produce fewer gestures due to not receiving any feedback. However, Bavelas and Healing (2013) suggested that addressees can become verbally responsive when behind a screen to compensate for not being able to produce visual responses . In the current study, we manipulated the addressee's verbal responses to establish if the addressee's behaviour could be responsible for any effect of visibility we find. The addressee was verbally responsive for half of the trials, and non-responsive for the other half. We also obtained ratings of the addressee's facial responses to establish if the addressee is more visually responsive in the responsive condition than the non-responsive condition which may be responsible for an effect on gesture (Bavelas & Healing, 2013). Following Beattie and Aboudan (1994), speakers should produce gestures at a higher rate in the responsive addressee condition than the non-responsive addressee condition.

If we find that speakers gesture more in the no screen and shoulder height screen conditions than in the full screen condition, it is still not clear whether visibility of the face is crucial, or visibility of any body part is sufficient, including the visibility of gestures (without visibility of the face). Experiment 3 investigated this possibility by comparing speaker's gesture production in two novel conditions as well as the no screen and full screen conditions. In a first novel condition, the speaker talked to the addressee with a chin upwards screen suspended between them, so the addressee could see the speaker's gestures, but the speaker and addressee could not see each other's faces. In a second novel condition, the speaker talked to the addressee behind a knee upwards screen suspended between them so the speaker and addressee could not see each other's faces or gestures, but the speaker would be able to see the addressee's shins. We predict that only visibility of the face is relevant for the effect of visibility on gesture, and that speakers should only produce more gestures when the addressee's face can be seen.

3.2. Experiment 1

3.2.1. Introduction

We investigated the co-speech gestures produced when speakers discussed abstract topics, namely, social dilemmas to determine if the speakers designed their gestures for the addressee via a perspective taking or a cue-based heuristic mechanism.

3.2.2. Methods

3*.2.2.1. Participants*

Participants were recruited in two ways. One set of participants were first year Psychology undergraduate students from the University of Warwick who took part for course credit. The other set of participants were volunteers (from both within and outside the

University) who were recruited using the department's participant recruitment system. Participants were only included in analysis if they spoke English as their first language. 38 native English-speakers participated in the experiment. Two participants were excluded from the analysis because they were holding an object during the task, and thus could not produce gestures. Five participants were excluded due to experimenter or video error. The data from the remaining 31 participants were analysed (Age, M = 25.23, SD = 9.67). An additional 23 participants (Psychology undergraduates) rated the addressee's visual responsiveness (Age, M = 21.61, SD = 7.35). The number of participants recruited was based off previous studies investigating the effect of visibility on gesture (e.g. Bavelas et al., 2008) in which 30 participants were analysed.

3*.2.2.2. Design*

The experiment used a 2x3 design with the two within-participant independent variables: visibility (no screen, shoulder height screen, full screen), addressee responsiveness (responsive, non-responsive). Three dependant variables were measured for the gesture production task: the rate of representational gestures per 100 words, the rate of interactive gestures per 100 words, the average representational gesture height. An additional dependent variable was measured in order to assess how much the addressee visually responded in the gesture production task: the addressee's visual responsiveness.

3*.2.2.3. Materials*

To elicit gesture, participants saw six social dilemmas and answered three questions for each dilemma. The questions were written underneath the dilemma for the participant to read. Four of the social dilemmas were used by Chu et al. (2014), and a further two were written in the same style for the current study. The full set of social dilemmas are in "Social Dilemmas" in the supplementary materials.

Example Social Dilemma:

"**Hannah** and **Louise** have been best friends for many years and have just started university together. **Hannah** is enjoying the freedom of university and particularly enjoys socialising. **Louise** is missing home and is not enjoying her university experience; she hasn't made any new friends and is very dependent on **Hannah**. **Hannah** feels guilty for having fun with her new friends knowing **Louise** is miserable. **Hannah** tries to encourage **Louise** to go to parties with her, but knows her new friends think **Louise** is geeky and do not like her. In recent weeks, **Hannah** has started to feel torn as both her new friends and **Louise** have been inviting her to go out increasingly often. She enjoys the company of her new friends but feels obliged to spend time with **Louise**.

- If you were in **Hannah**'s situation, what would you do and why?
- What do you think **Louise** would feel about your decision?
- What do you think your new friends would feel about your decision?"

Visibility was manipulated by adjusting a screen fixed between two height-adjustable tripods. See Figure 1

Figure 1

The experimental set up used for Experiments 1, 2, and 3. The participant sat on the chair on the right bottom, and the addressee (i.e., the experimenter) sat on the other side of the screen. The laptop is situated next to the participant. One camera recorded the participant, and a second camera recorded the experimenter. The top of the screen was adjusted for different conditions by changing the height of the tripod, while the bottom of the screen was kept 10cm above the floor.

Two Canon Legria R66 video cameras were used to record during the experiment. The recordings of the experimenter were used as stimuli for participants to provide the addressee's visual responsiveness rating.

3*.2.2.4. Gesture Elicitation Procedure*

The experimenter informed the participants that they would be taking part in an experiment about social dilemmas. The experimenter did not inform the participants that the study was investigating gesture. Participants were seated opposite the experimenter's chair. The participants took the role of speaker, and the experimenter took on the role of addressee. The stimuli were displayed to the side of the participants on a laptop.

The participant read the first social dilemma and answered the questions about the dilemma to the experimenter. There was no time limit and the dilemma and questions were visible to the participant as they answered. When the participant had finished answering, the participant moved the PowerPoint slide on to the next social dilemma.

In the *no screen* condition, the screen was set aside so that the participant and experimenter could both see each other with no barrier. In the *shoulder height screen* condition the screen was placed between the experimenter and participant, and the top of the screen was raised to the shoulder height so that they could only see each other's chins and above. In the *full screen* condition, the screen was placed between the experimenter and participant, and the top of the screen was raised to the height such that they could not each other at all.

In the *addressee responsive* condition, the experimenter produced generic responses such as "yeah, okay, uh-huh and mm-hmm" whenever appropriate. The experimenter also produced specific responses that fit into the context. The experimenter used statements such as "Ah I get what you mean" in response to a difficult to explain piece of description. In the *addressee non-responsive* condition, the experimenter was silent throughout.

The participants read each dilemma and answered its associated questions in turn to the experimenter. Participants did this for all six social dilemmas. We counterbalanced condition order using the Latin square technique. The order for the first participant was: *no screen responsive, no screen non-responsive, shoulder height screen responsive, shoulder height screen non-responsive, full screen responsive and full screen non-responsive*. The order of conditions was rotated between participants. Thus, each condition appeared equally often in each trial (trial one to trial six). The six social dilemmas were presented in a fixed order.

Participants were then asked to fill in a short questionnaire to obtain handedness information and biographical information as well as a full video consent form, Participants were then fully debriefed.

3*.2.2.5. Addressee Responsiveness Procedure*

A new group of participants were asked to rate the addressee's (i.e., the experimenter's) visual responses with facial expression and head movements in the videos obtained in the gesture elicitation part of the experiment. The videos were cut down into clips with each clip showing the addressee in a single trial. The participant provided one rating for each trial using a scale from 1 to 7. Participants were instructed that "a rating of 1 indicated no responsiveness whatsoever and a rating of 7 indicated that the addressee responded very strongly to the speaker with their facial expression and head movements". Participants were also instructed to not take the jaw movements resulting from speaking into account, but changes in facial expression or head movements during speech should be counted.

The participants saw the clips from the conditions (trials) where the addressee's face could be seen by the speaker: *no screen responsive, no screen non-responsive, shoulder height screen responsive, shoulder height screen non-responsive.* The participants saw clips of the addressee, taken from the sessions with four speakers. As each speaker had four conditions, the participants rated 16 clips in total. The conditions were presented to participants in this fixed order, starting with a randomly selection condition (e.g., if *"shoulder height screen responsive"* was randomly selected to be first, the order the conditions would be shown in would be: *shoulder height screen responsive, shoulder height screen non-responsive, no screen responsive, no screen non-responsive).* Each condition was selected to be first an equal number of times. The videos were shown with audio, as the audio was necessary for participants to establish when a behaviour was in response to something said by the speaker.

3*.2.2.6. Gesture Coding*

Participants' gestures were coded using a modified version of the coding manual used by Chu et al. (2014). Participants' gestures were coded during the stroke phase of the preparation-stroke-hold-retract breakdown of gestures (Kita et al., 1997). If there was a clear preparation or hold between two strokes, they were coded as separate gestures. Gestures were categorised based of their form and function for analysis. The two key gesture types were representational gestures and interactive gestures. Representational gesture and interactive gesture each consisted of several sub-categories of gesture. The coding criteria are in "Gesture Coding Manual" in the supplementary materials.

Representational gestures are made up of depictive, deictic and conduit gestures. Depictive gestures are movements which can depict the actions of an agent, the movement of an object or a property of either an actor or agent. These are made up of two sub-categories. Iconic gestures (McNeill, 1992) depict some physical action, object, or property. Depictive gestures can also represent a metaphorical concept or property and are called metaphorical gestures (McNeill, 1992). Deictic gestures are pointing gestures which identify a location in the space. Some representational gestures are classified as representationally unclear if they have no clear meaning but still seem to be shaped to represent something. Conduit gestures (McNeill, 1992) are produced by the speaker moving an open palm towards the addressee to present an idea. The speaker should be certain about what they are saying. Conduit gestures can be identified by a clear forward movement of the hand toward the addressee.

Interactive gestures (Bavelas et al., 1992) are made up of listener indexing, conduit and palm revealing emblem gestures. Listener-Indexing gestures refer to the addressee. A gesture is coded as a listener-indexing gesture when the hand moving towards the addressee and accompany speech that references the addressee. The speaker would typically have to be making eye contact with the addressee in previous research (Chu et al., 2014), but this requirement was dropped because in the full screen condition, the speaker and addressee could not make eye contact. Conduit gestures (as seen in the previous paragraph) are coded as interactive gestures as well as representational gestures. This is because conduit gestures directly index the addressee, as the idea being presented is being placed on the palm and moved toward the addressee. Palm revealing gestures included shrugs with the palms out and facing upwards to express empty handedness. The message of uncertainty, or a lack of knowledge to share should be conveyed.

Interactive gestures were not analysed because they were produced rarely in our data set. Only 23 out of the 31 participants produced interactive gestures, and only 1 participant produced interactive gestures in every condition. On average each participant only produced 4.71 interactive gestures across all conditions (SD = 8.37/SE = 1.5). Interactive gestures were even less frequent in Experiments 2 and 3, and thus they were not analysed (see supplementary materials for the descriptive statistics).

Other gesture classifications were coded, but not included in the calculation of dependant variables in the current study. These included beat gestures which are small movement up and down, typically in the lap and the hand shape is loose. There should be no clear meaning assignable to the gesture. Emblems are symbolic gestures where a hand movement has a conventional meaning. An example of this being a thumbs-up emblem. Metacognitive gestures are repetitive movement that indicate thinking. Generally, this involves repeated finger tapping or rotating the hand at the rest while pausing speech to search for a word or phrase. Unclear gestures are those that fail to fit into any of these categories.

Representational gestures were coded into three different heights. The first height category was "lower than waist". The waist was defined as being part way between a person's hips and the bottom of their ribcage. The second height category was "between waist and chin". The third height category was "above chin". For any given gesture, an imaginary line is drawn across the participant across their waist and another line is drawn across the base of the chin. The lines were drawn parallel to horizontal shelves in the background to account for the angle of the camera. The height of a gesture was determined by the highest point of the hand or figure during the gesture movement. The location of the highest point is compared to the waist and chin lines to determine the height coding. Lower than waist gestures were given a value of 1. Between waist and chin gestures were given a value of 2. Above chin gestures were given a value of 3.

3.*2.2.7. Inter-Rater Reliability*

A first coder initially coded all of the participant's gestures. To establish reliability a second and third coder independently assessed 20% of the data. To establish gesture rate reliability, the second coder coded one condition from 24 participants. There was a strong positive correlation between the first and second coder, $r(22) = .979$, $p < .001$. To establish representational versus non-representational gesture classification and height reliability, the third coder coded all the data from 8 participants. Agreement between the first and third coder was 92% ($N = 326$) when determining if a gesture was representational or not. Agreement between the first and third coder for gesture height was 79% ($N = 326$) for determining the height of representational gestures.

3.*2.2.8. Transparency and openness*

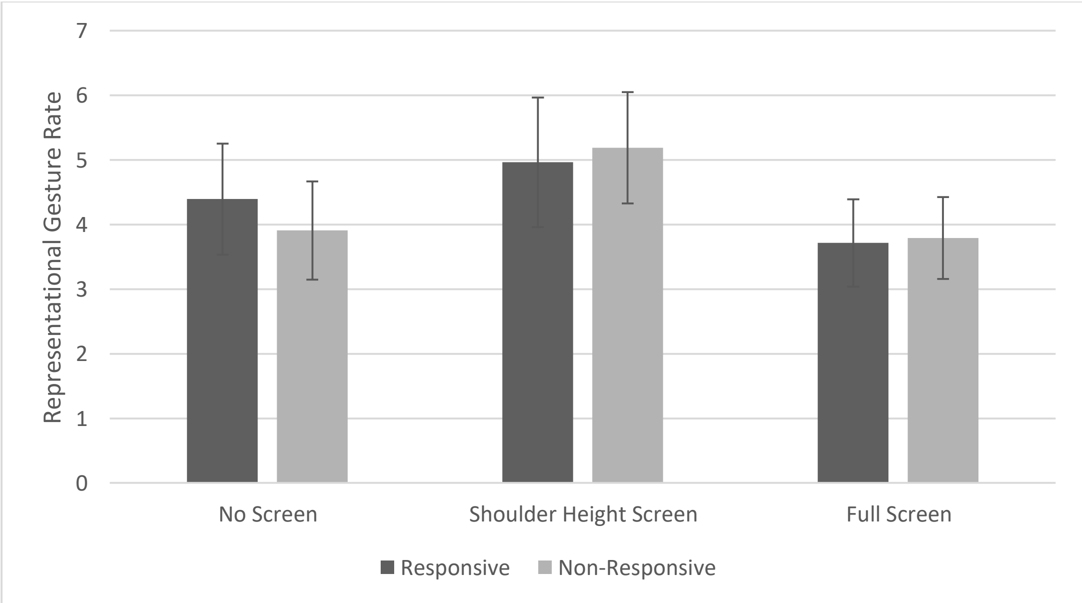
We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and research materials are available at the following link: https://osf.io/pbnfd/?view_only=b9a69f2fbbd843b6a6a3b7a30fbcdbbf. Data were analysed using SPSS version 27.0. This study's design and its analysis were not pre-registered.

3.2.3. Results

The representational gesture rate was analysed in a 2x3 analysis of variance with visibility (no screen, shoulder height screen, full screen) and addressee responsiveness (responsive, non-responsive) as within-participant independent variables (See Figure 2 for descriptive statistics). There were no statistically significant main effects of visibility, $F(2, 60) = 2.64$, $p = .080$, $\eta_p^2 = .081$, and addressee responsiveness, $F(1, 30) = 0.02$, p = .890, $\eta_p^2 = .001$, on representational gesture rate. There was no statistically significant interaction effect between visibility and addressee responsiveness on representational gesture rate, $F(2, 60) = 0.31$, $p = .733$, $\eta_p^2 = .010$.

Figure 2

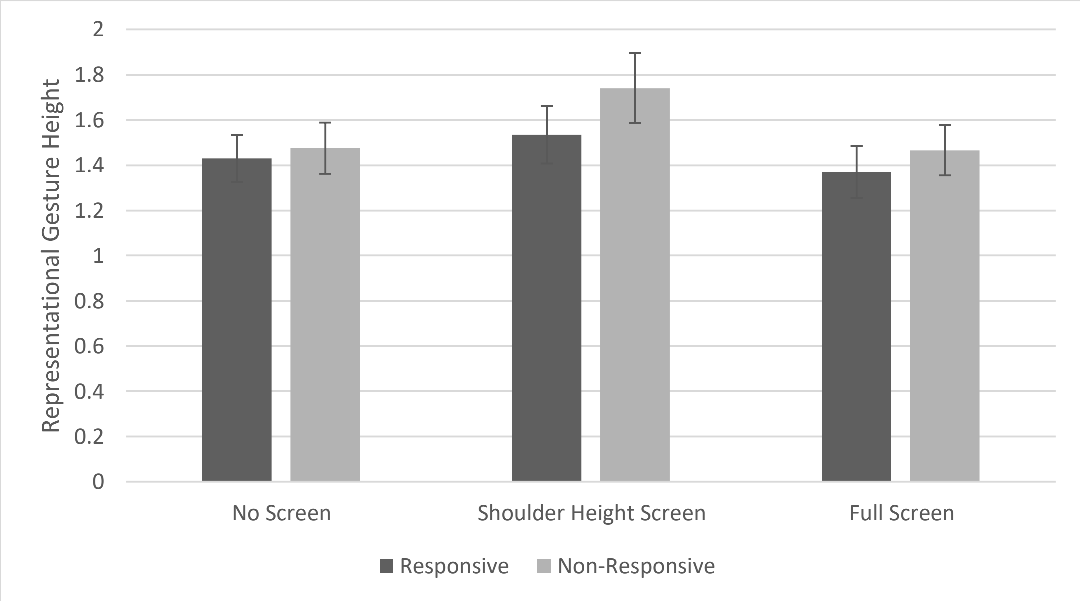
Mean representational gesture rate (per 100 words) in the no screen, shoulder height screen, and full screen conditions when the addressee was responsive and non-responsive in Experiment 1 (social dilemma explanation task). The error bars represent standard errors.



We calculated the mean height of representational gestures for each condition for each participant. The mean representational gesture height was analysed in a 2x3 analysis of variance with visibility (no screen, shoulder height screen, full screen) and addressee responsiveness (responsive, non-responsive) as within-participant independent variables (See Figure 3 for descriptive statistics). There were no statistically significant main effects of visibility, $F(2, 30) = 2.63$, $p = .089$, $\eta_p^2 = .149$, and addressee responsiveness, $F(1, 15) = 2.69$, $p = .122$, $\eta_p^2 = .152$. There was no statistically significant interaction effect between visibility and addressee responsiveness, $F(2, 30) = 0.60$, $p = .554$, $\eta_p^2 = .039$.

Figure 3

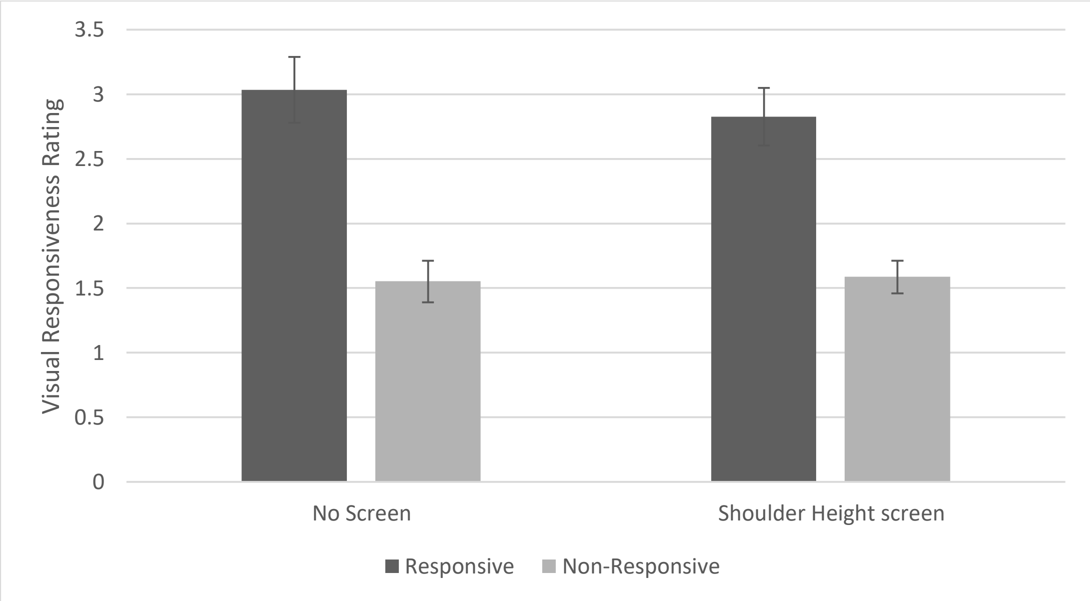
Mean representational gesture height in the no screen, shoulder height screen, and full screen conditions when the addressee was responsive and non-responsive in Experiment 1 (social dilemma explanation task). The error bars represent standard errors.



The addressee's visual responsiveness (i.e., the face and the head movements) was analysed in a 2x2analysis of variance with visibility (no screen, shoulder height screen) and addressee responsiveness (responsive, non-responsive) as within-participant independent variables (See Figure 4 for descriptive statistics). There was no statistically significant main effect of visibility, $F(1, 28) = 0.305$, $p = .583$, $\eta_p^2 = .011$. There was a statistically significant main effect of addressee responsiveness, $F(1, 28) = 35.91$, $p < .001$, $\eta_p^2 = .562$, with the addressee being rated as more visually responsive in the responsive condition than the non-responsive condition. There was no significant interaction effect between visibility and addressee responsiveness, $F(1, 28) = 1.04$, $p = .316$, $\eta_p^2 = .036$.

Figure 4

Addressee's visual responsiveness in the no screen and shoulder height screen conditions when the addressee was responsive and non-responsive in Experiment 1 (social dilemma explanation task). The error bars represent standard errors.



3.2.4. Discussion

There were no significant effects of visibility or addressee responsiveness on the representational gesture rate. The findings did not support either the perspective taking or the cue-based heuristic view of the effect of visibility on gesture. We cannot conclude that either a heuristic mechanism or a perspective taking mechanism is responsible for the effect of visibility on gesture. It was found that the addressee (the experimenter) was more visually responsive (with his head and face) in the responsive condition than in the non-responsive condition. The addressee manipulated verbal responses and did not deliberately change the visual responsiveness between the conditions. However, visual responses came naturally with verbal responses. This experiment did not provide evidence for the claim (Bavelas & Healing, 2013) that people produce more gestures when the addressee provide more verbal/visual feedback.

Experiment 1 may have failed to find the effect of visibility on gesture because the effect only occurs when speakers are trying to convey dynamic spatial information because gesture is communicatively effective in these domains (Hostetter, 2011). The speaker may not have produced more gestures when the addressee could see the speaker's gestures because gesture would not be as useful for communication during the task.

To test our predictions that speakers' design their gestural communication for the addressee via a cue-based heuristic, we need to elicit the effect of visibility on gesture. In Experiment 2 we will run the study again but use different stimuli to elicit more gestures that convey dynamic spatial information from the speaker rather than abstract gestures.

3.3. Experiment 2

3.3.1. Introduction

In Experiment 2 we again aimed to determine if the speakers designed their gestures for the addressee via a perspective taking or a cue-based heuristic mechanism. Experiment 2 was carried out in the same way as Experiment 1, except for the stimuli to elicit gestures. We used cartoon clips that have been used in previous studies that did find the effect of visibility on gesture. Cartoon segments have been used to elicit the effect of visibility on gesture before (Alibali et al., 2001; Mol et al., 2009a, 2009b). The cartoon stimuli should elicit gestures conveying information about the characters relative locations and movements (Dynamic spatial information), and actions (motoric information). If gestures conveying spatial or motoric information result in gesture becoming more useful for communication (Hostetter, 2011), speakers should include gestures when designing their communication for the addressee and produce more gestures when speakers believe they will benefit the addressee. Our predictions for Experiment 2 remain the same as the predictions for Experiment 1.

To ensure that speakers were not using their heads to gesture more when their hands were hidden behind the shoulder height screen than when their hands were visible to the speaker, we measured speaker's representational head movements.

We also carried out an exploratory analysis comparing the results of Experiments 1 and 2. We predict that speakers should produce more representational gestures in Experiment 2 (Dynamic spatial stimuli) than in Experiment 1 (abstract stimuli). We also predict that there should be a significant interaction between the experiment and gesture visibility, indicating that the effect of visibility on gesture occurred in Experiment 2 but not Experiment 1.

3.3.2. Methods

3.*3.2.1. Participants*

Participants were all first year Psychology undergraduate students from the University of Warwick. Participants were only included in analysis if they spoke English as their first language. 34 native English-speakers participated for course credit. Two participants were excluded from the analysis because they were holding an object during the task, and thus could not produce gestures. One participant was excluded due to video error. The data from the remaining 31 participants were analysed (Age, *M* = 18.48, *SD* = 0.68). An additional 11 participants (Psychology undergraduates) rated the addressee's visual responsiveness (Age, *M* = 19, *SD* = 0.89). The number of participants recruited in Experiment 2 was based on the number of participants recruited in Experiment 1.

3.*3.2.2. Design*

The design of Experiment 2 was similar to the design of Experiment 1 with the exception of measuring a new dependant variable: Representational head movement rate (representational head movements per 100 words).

3.*3.2.3. Materials*

The screen was the same as in Experiment 1. Participant's speech and gesture was elicited by video clips of six episodes from a Tweety and Sylvester film entitled "Canary Row" (which was also used in McNeill, 1992).

3.*3.2.4. Gesture Elicitation procedure*

The procedure was the same as Experiment 1 except that participants were shown cartoon clips to elicit speech and gesture.

The participants were shown how to start each video clip and move on to the next. The laptop was placed on a table to the side of the participant. Participants used headphones to listen to the audio. Participants were instructed to watch the first video clip twice. Participants then placed the headphones back on the table. Participants were instructed to turn to face the experimenter and to describe the cartoon in as much detail as possible. Participants were prompted to go on to the next cartoon by the experimenter. This repeated for all six conditions.

Counterbalancing was the same as Experiment 1. Gesture visibility and addressee responsiveness were both manipulated in the same way as Experiment 1.

3.*3.2.5. Addressee Responsiveness Procedure*

Addressee responsiveness ratings were obtained in the same way as in Experiment 1.

3.*3.2.6. Gesture Coding*

Hand gestures were coded in the same way as Experiment 1. Representational head movements were coded using a supplement to the coding manual used in Experiment 1. Participants head movement gestures were coded during the stroke phase of the preparation-stroke-hold-retract breakdown of gestures (Kita et al., 1997). If there was a clear preparation or hold between two strokes, they were coded as separate head movement gestures. Gestures were categorised based of their form and function for analysis. Representational head movements were defined as movements of the head which can be interpreted as depicting some physical action, property, movement of a character or object, metaphorical concept, or identifying a location in space. Only representational head movements were coded for the current study. The coding supplement for head movement gestures can be found in "Head Movement Coding Supplement" in the supplementary materials.

3.*3.2.7. Inter-Rater Reliability*

A first coder initially coded all of the participant's gestures. To establish reliability a second, third, and fourth coder independently assessed 20% of the data. To establish gesture rate reliability, the second coder coded one condition from 24 participants. There was a strong positive correlation between the first and second coder, $r(22) = .974$, $p < .001$. To establish representational versus non-representational gesture classification and height reliability, the third coder coded all the data from 8 participants. Agreement between the first and third coders was 91% (N = 1,224) when determining if a gesture was representational or not. Agreement between the first and third coder was 87% (N = 1,224) for determining the height of representational gestures. To establish representational head movement rate reliability, the fourth coder coded two conditions from all participants. The rates of representational head movements (per 100 words) based on the first and fourth coders' coding positively correlated with each other, $r(62) = .756$, $p < .001$.

3.*3.2.8. Transparency and openness*

Experiment 2 adhered to the same transparency and openness declarations as Experiment 1.

3.3.3. Results

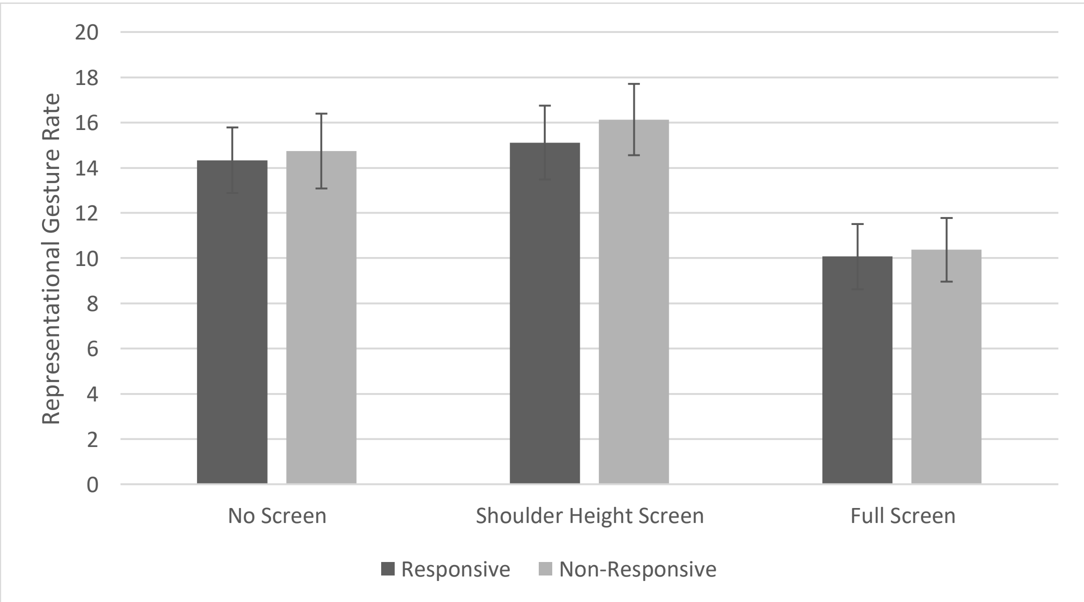
The representational gesture rate was analysed in a 2x3 analysis of variance with visibility (no screen, shoulder height screen, full screen) and addressee responsiveness (responsive, non-responsive) as within-participant independent variables (See Figure 5 for descriptive statistics). There was a statistically significant main effect of visibility, (Greenhouse-Geisser corrected degrees of freedom) $F(1.35, 40.39) = 15.28$, $p < .001$, η_p^2 = .337, but no significant main effect of addressee responsiveness, $F(1, 30) = 0.63$, p = .433, η_p^2

= .021. There was no statistically significant interaction effect between visibility and addressee responsiveness, $F(2, 60) = 0.13$, $p = .881$, $\eta_p^2 = .004$.

To further explore the main effect of visibility, we ran posthoc pair-wise comparisons (as advised by Howell (2017) for a comparison of three means, we used Fisher's LSD test). Speakers produced representational gestures at a significantly higher rate in the no screen than the full screen condition $t(30) = 3.50$, $p = .001$, $d = .629$, and a significantly higher rate in the shoulder height screen condition than the full screen condition, $t(30) = 4.64$, $p < .001$, $d = .833$. Speakers did not produce representational gestures at a significantly higher rate in the no screen than the shoulder height screen conditions $t(30) = -1.90$, $p = .067$, $d = -.341$.

Figure 5

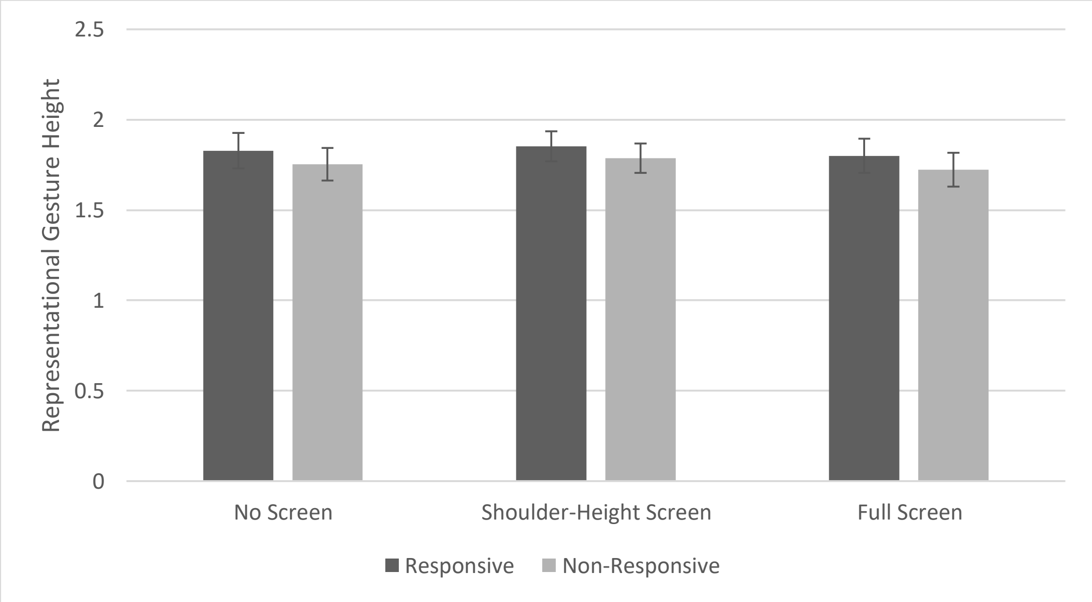
Mean representational gesture rate (per 100 words) in the no screen, shoulder height screen, and full screen conditions when the addressee was responsive and non-responsive in Experiment 2 (Cartoon description task). The error bars represent standard errors.



The mean representational gesture height was analysed in a 2x3 analysis of variance with visibility (no screen, shoulder height screen, full screen) and addressee responsiveness (responsive, non-responsive) as within-participant independent variables (See Figure 6 for descriptive statistics). There were no statistically significant main effects of visibility, $F(2, 50) = 0.56$, $p = .575$, $\eta_p^2 = .022$, and addressee responsiveness, $F(1, 25) = 2.03$, $p = .167$, $\eta_p^2 = .075$. There was no statistically significant interaction effect between visibility and addressee responsiveness, $F(2, 50) = 0.02$, $p = .981$, $\eta_p^2 = .001$.

Figure 6

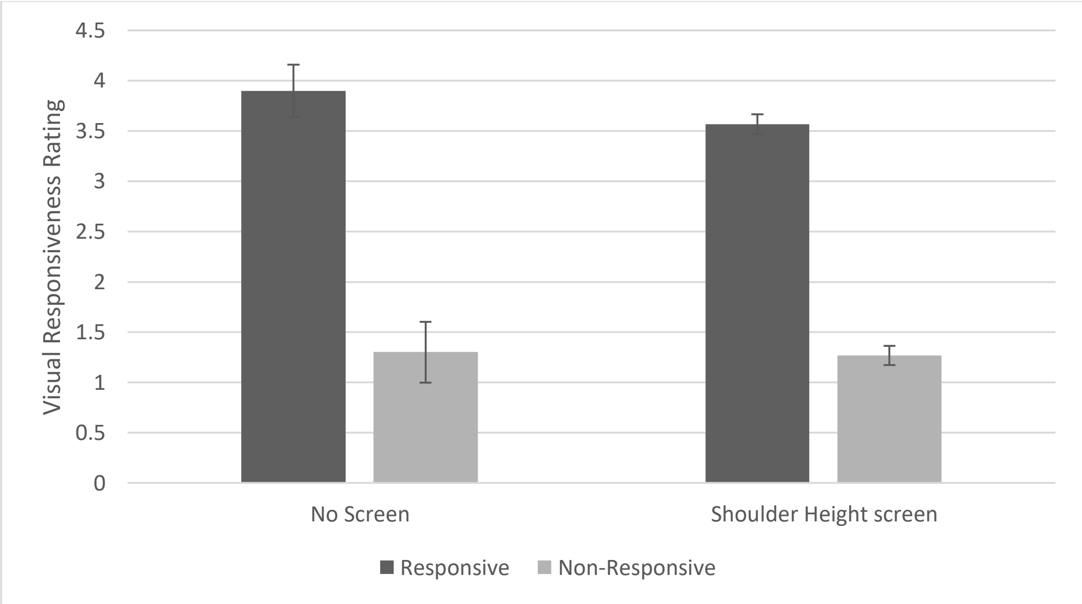
The mean representational gesture height in the no screen, shoulder height screen, and full screen conditions when the addressee was responsive and non-responsive in Experiment 2 (Cartoon description task). The error bars represent standard errors.



The addressee's visual responsiveness (with the facial and head movements) was analysed in a 2x2 analysis of variance with visibility (no screen, shoulder height screen) and addressee responsiveness (responsive, non-responsive) as within-participant independent variables (See Figure 7 for descriptive statistics). There was no statistically significant main effects of, $F(1, 29) = 2.722$, $p = .110$, $\eta_p^2 = .086$. There was a statistically significant main effect of addressee responsiveness, $F(1, 29) = 105.13$, $p < .001$, $\eta_p^2 = .784$, with the addressee being rated as more visually responsive in the responsive condition than the non-responsive condition. There was no statistically significant interaction effect between visibility and addressee responsiveness, $F(1, 29) = 2.157$, $p = .153$, $\eta_p^2 = .069$.

Figure 7

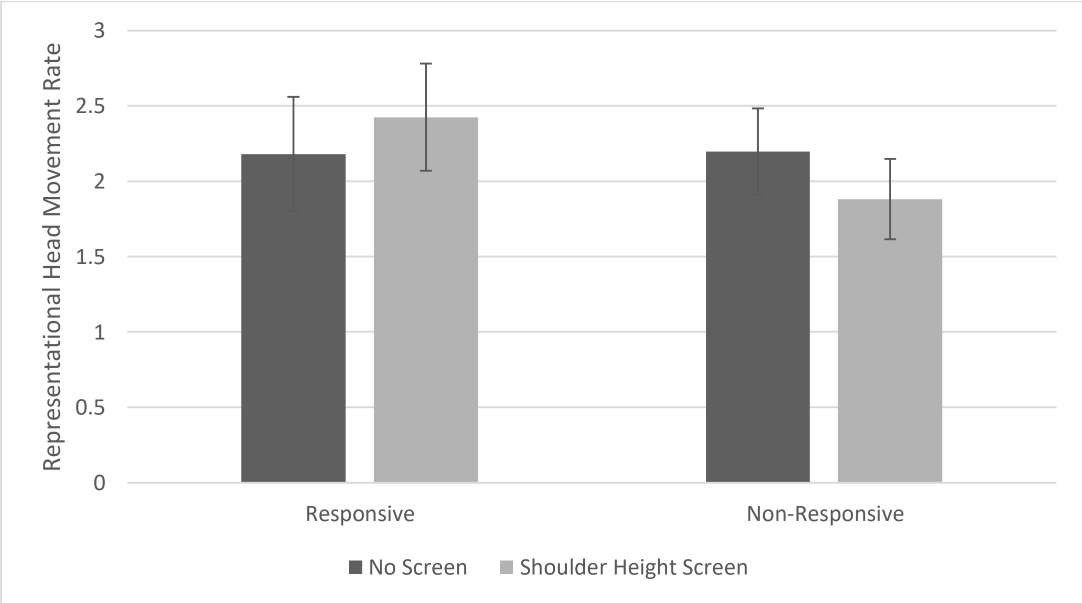
Addressee's visual responsiveness in the no screen and shoulder height screen conditions when the addressee was responsive and non-responsive in Experiment 2 (Cartoon description task). The error bars represent standard errors.



The representational head movement rate was analysed in a 2x2 analysis of variance with visibility (no screen, shoulder height screen) and addressee responsiveness (responsive, non-responsive) as within-participant independent variables (See Figure 8 for descriptive statistics). There were no statistically significant main effects of visibility, $F(1, 30) = 0.032$, $p = .860$, $\eta_p^2 = .001$, and addressee responsiveness, $F(1, 30) = 1.143$, $p = .293$, $\eta_p^2 = .037$, on representational gesture rate. There was no statistically significant interaction effect between visibility and addressee responsiveness on representational gesture rate, $F(1, 30) = 1.124$, $p = .298$, $\eta_p^2 = .036$.

Figure 8

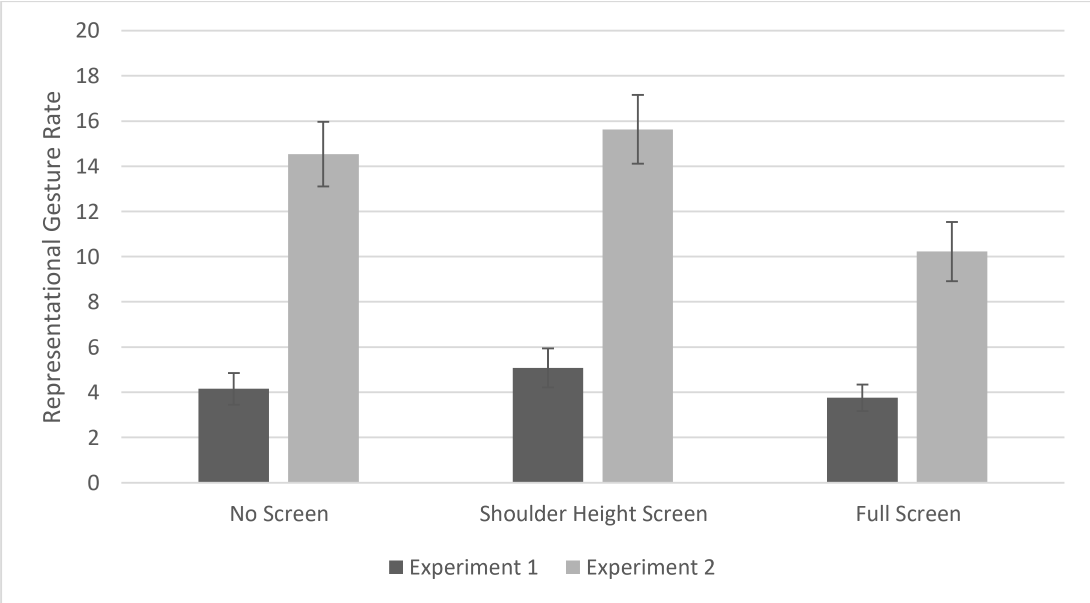
Mean representational head movement rate (per 100 words) in the no screen and shoulder height screen conditions when the addressee was responsive and non-responsive in Experiment 2 (Cartoon description task). The error bars represent standard errors.



The representational gesture rates from Experiment 1 and 2 were analysed in a 2x3 mixed factor analysis of variance with visibility (no screen, shoulder height screen, full screen) as a within-participant independent variable, and Experiment (Experiment 1, Experiment 2) as a between-participant independent variable (See Figure 9 for descriptive statistics). There was a statistically significant main effect of visibility, (Greenhouse-Geisser corrected degrees of freedom) $F(1.67, 100.46) = 16.8$, $p < .001$, $\eta_p^2 = .219$, and a statistically significant main effect of experiment, $F(1, 60) = 40.15$, $p < .001$, $\eta_p^2 = .401$. There was a statistically significant interaction effect between visibility and experiment, $F(2, 120) = 7.52$, $p = .001$, $\eta_p^2 = .111$. Visibility influenced the representational gesture rate in Experiment 2 but not in Experiment 1.

Figure 9

Mean representational gesture rate (per 100 words) in the no screen, shoulder height screen, and full screen conditions in both Experiment 1 (social dilemma explanation task) and Experiment 2 (Cartoon description task). The error bars represent standard errors.



3.3.4. Discussion

There are three key findings. First, speakers' representational gesture rates were higher in the no screen and shoulder height screen conditions than in the full screen condition. Second, speakers did not produce their gestures at different heights in any of the gesture visibility conditions. Third, speakers did not produce significantly more representational gestures in the addressee responsive condition than the addressee non-responsive condition. The findings indicated that speakers produced more representational gestures in the conditions when the speaker and addressee could see each other's faces, rather than when the addressee could see the speaker's gestures. That is, the results provided unequivocal evidence for a cue-based heuristic for audience design behaviour in gesture.

The findings provided no evidence to support the perspective taking view that speakers would only gesture more when the addressee can see the speaker's gestures. Speakers did not produce more gestures when their gestures could be seen by the addressee, and speakers did not produce their gestures above the shoulder height screen so the addressee could see the gestures. Furthermore, there is no evidence to suggest speakers compensated for their gestures not being visible by producing more head movements as speakers did not produce more representational head movements in the shoulder height screen condition than in the no screen condition

The findings provided no evidence to suggest that the effect of visibility was due to the confounding of visibility and addressee responsiveness, as suggested by Bavelas and Healing (2013). First, there was no effect of the addressee's responsiveness on the speaker's representational gesture rate. As with Experiment 1, addressee's facial expressions were rated as more responsive in the responsive condition than the non-responsive condition. Second, the addressees' verbal responsiveness was manipulated in the same way in Experiments 1 and 2, and this manipulation led to the same pattern of addressee's visual responsiveness in Experiments 1 and 2. Yet, the visibility manipulation influenced the gesture rate only in Experiment 1. Together, these findings indicate that the visibility effect cannot be attributed to patterns of the addressee responsiveness.

While Experiment 2 suggests the speaker seeing the addressee faces triggers the cue-based heuristic, it is not clear if visibility of other body parts may trigger the heuristic. The findings of Experiment 2 do not make clear if visibility of the face is the only cue responsible for an effect of visibility on gesture. It is possible that there are multiple cues that could gesture production. For instance, visibility of the speaker's gestures can act as a second cue and increase gesture production even when the addressee's face is not visible. In addition, it is possible that the speaker seeing any part of the addressee could act as a cue to the heuristic, not just the face. Experiment 3 investigated these possibilities.

3.4. Experiment 3

3.4.1. Introduction

There are two alternative ways in which the effect of visibility on gesture production could occur. First, The speaker's gestures being visible to the adressee could trigger the effect of visibility on gesture even when the addressee's face is not visible to act as a cue. This would be in line with the claim in previous research (e.g. Alibali et al., 2001). We tested this hypothesis by using the screen to prevent the speaker and addressee from seeing each other from the chin upwards. The speaker and addressee were not able to see each other's faces, but the speaker's gesture was clearly visible.

Second, it is possible that any part of the addressee (a visual indicator of the addressee's presence) can act as a cue for a heuristic mechanism. If so, then when the speaker sees the addressee's lower legs and feet without seeing the speaker's gestures or face, the heuristic should be triggered. We tested this hypothesis by using the screen to prevent the speaker and addressee from seeing each other from the knees upwards. The addressee was not able to see either the speaker's face or gestures, but the speaker was still able to see the addressee's lower legs (a visual indicator of the addressee's presence).

Similar to Experiments 1 and 2, we used a no screen condition and a full screen condition to compare to the new chin upwards and knee upwards conditions.

We hypothesised that the speaker seeing the addressee's face uniquely triggers the heuristic. Therefore, speakers should produce more gestures when the speaker and addressee can see each other's faces than when they cannot even if the speaker and addressee can see each other's gestures or shins. Thus, we predict that speakers should produce more representational in the no screen condition than the chin upwards, knee upwards and full screen conditions.

3.4.2. Methods

3.*4.2.1. Participants*

Participants were all first year Psychology undergraduate students from the University of Warwick. Participants were only included in analysis if they spoke English as their first language. 34 native English-speakers participated for course credit. One participant was excluded from the analysis because they failed to follow instructions. The data from the remaining 33 participants were analysed (Age, *M* = 18.4, *SD* = 0.56). The number of participants recruited for Experiment 3 was based on the number of participants recruited in Experiment 1.

3.*4.2.2. Design*

Experiment 3 used a design with one within-participant independent variable: visibility (*no screen, chip upwards screen, knee upwards screen, full screen*). Three dependant variables were measured: the rate of representational gestures per 100 words, the mean representational gesture height, and the addressee's visual responsiveness.

3.*4.2.3. Materials*

The apparatus was the same as in Experiment 2. Participant's speech and gesture was elicited by four video clips from the six used in Experiment 2.

3.*4.2.4. Procedure*

The procedure was the same as in Experiment 2 except that participants only described four clips, rather than six. Counterbalancing was done in the say way as Experiments 1 and 2, with the initial order being: *no screen, chip upwards screen, knee upwards screen, full screen.*

In the *no screen* condition, the screen was placed aside so that the speaker and addressee could both see each other with no barrier. In the *chin upwards screen* condition, the screen was placed between the experimenter and participant so that they could only see each other from the chin downward but could not see each other's faces. The bottom of the screen was elevated on the tripods so the bottom was at chin height between the participant and experimenter. In the *knee upwards screen* condition, the screen was placed between the experimenter and participant so that they could only see each other from the knee downwards, but could not see each other's upper legs, torso or face. The screen was elevated on the tripods so the bottom of the screen was at knee height between the participant and experimenter. In the *full screen* condition, the screen was placed between the experimenter and participant so that they could not see any of each other.

3.*4.2.5. Gesture Coding*

Gesture coding was the same as in Experiments 1 and 2.

3.*4.2.6. Inter-Rater Reliability*

A first coder initially coded all the participant's gestures. To establish reliability a second and third coder independently assessed 20% of the data. To establish gesture rate reliability, the second coder coded one condition from 24 participants. There was a strong positive correlation between the first and second coder, $r(24) = .931$, $p < .001$. To establish representational versus non-representational gesture classification and height reliability, the third coder coded all the data from 8 participants. Agreement between the first and third coder was 87% ($N = 472$) when determining if a gesture was representational or not. Agreement between the first and third coder was 94% ($N = 472$) for determining the height of representational gestures.

3.*4.2.7. Transparency and openness*

Experiment 3 adhered to the same transparency and openness declarations as Experiment 1.

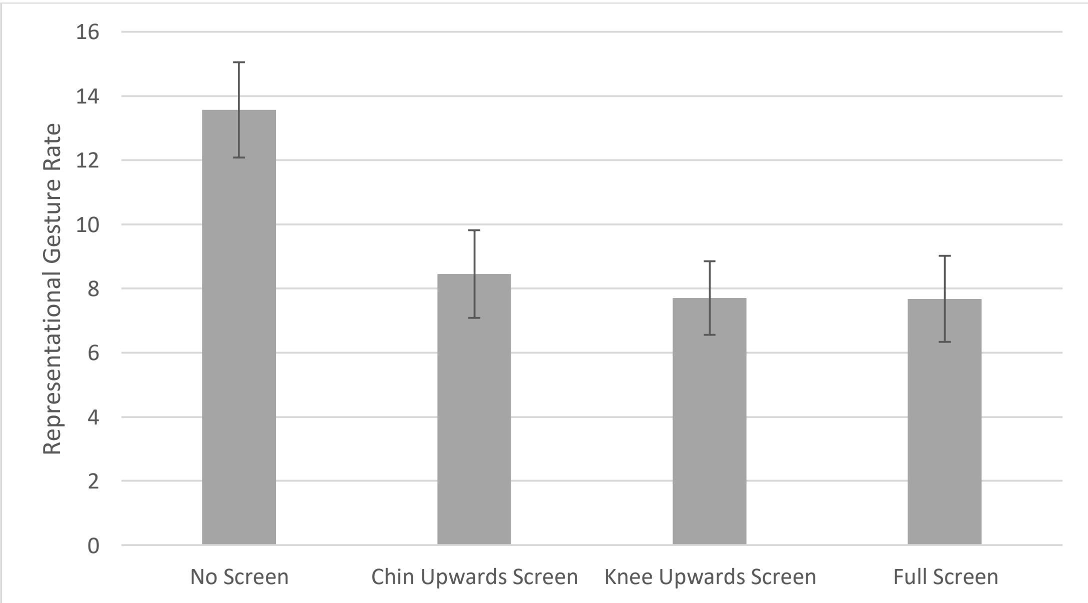
3.4.3. Results

The representational gesture rate was analysed in an analysis of variance with visibility (no screen, chin upwards screen, knee upwards screen, full screen) as the within-participant independent variable (See Figure 10 for descriptive statistics). There was a statistically significant main effect of visibility, $F(3, 96) = 13.24$, $p < .001$, $\eta_p^2 = .293$.

To further explore the main effect of visibility, we ran posthoc pair-wise comparisons with Bonferroni correction, with the alpha level of .017. Speakers produced representational gestures at a significantly higher rate in the no screen than the chin upwards screen condition, $t(32) = 4.35$, $p < .001$, $d = .757$, the knee upwards screen condition, $t(32) = 5.22$ $p < .001$, $d = .909$, and the full screen condition $t(32) = 4.66$, $p < .001$, $d = .811$. There were no differences between any of the other conditions (p >.48, $d < .124$).

Figure 10

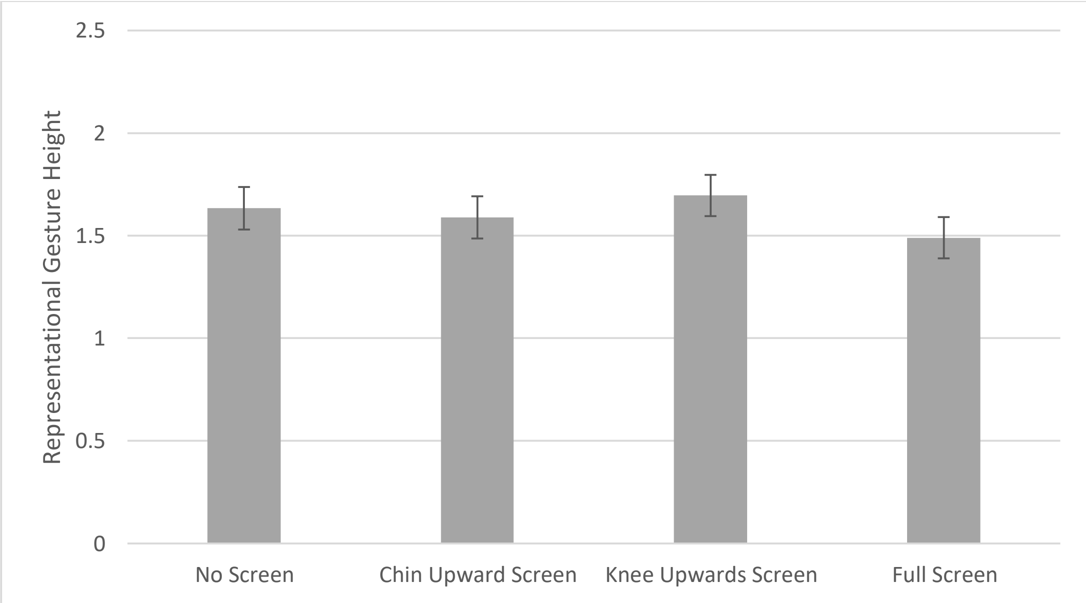
Mean representational gesture rate (per 100 words) in the no screen, chin upwards screen, knee upwards screen, and full screen conditions in Experiment 3 (Cartoon description task task). The error bars represent standard errors.



The mean representational gesture height was analysed in an analysis of variance with visibility (no screen, chin upwards screen, knee upwards screen, full screen) as the within-participant independent variable (See Figure 11 for descriptive statistics). There was no main effect of visibility, (Greenhouse-Geisser corrected degrees of freedom) $F(2.35, 53.95) = 1.84$, $p = .162$, $\eta_p^2 = .074$.

Figure 11

Mean representational gesture height in the no screen, chin upwards screen, knee upwards screen, and full screen conditions in Experiment 3 (Cartoon description task task). The error bars represent standard errors.



3.4.4. Discussion

There were two key findings. First, speakers produced more representational gestures in the no screen condition than in the chin upwards, knee upwards and full screen conditions. Second, speakers did not produce their gestures at different heights in any of the gesture visibility conditions. Speakers produced more representational gestures in the condition when the speakers and addressees could see each other's faces than in the conditions when the speaker and addressee could not see each other's faces. There is no evidence to suggest that the speaker's gesture being visible acts as a cue which triggers the speaker to produce more gestures. Furthermore, there is no evidence to suggest that any visual indicators of the addressee's presence can act as a cue which triggers the speaker to produce more gestures.

3.5. General Discussion

The current study investigated whether a cue-based heuristic mechanism can be responsible for audience design effects on co-speech representational gestures. In both Experiments 2 and 3, speakers produced more gestures when the addressee's face was visible to the speaker than when the addressee's face was not visible. Speakers did not raise their gestures around the shoulder height screen (Experiment 2) or lower their gestures below the chin upwards screen (Experiment 3) so the addressee could see the gestures.

Furthermore, it was not the case that speakers produced more gestures when the addressee could see the speaker's gestures than when the addressee could not. This was the case in the situations where the speaker can see the addressee's face (Experiment 2) or not (Experiment 3). Speaker's gesture production was unaffected by the addressee's responsiveness behaviour, suggesting that the effect of visibility in the current study was *not* caused by the disparity in responsiveness between the no screen and full screen conditions. Furthermore, an account based on the addressee's responsiveness cannot explain why the visibility effect was observed when the speech content was spatial (Experiment 2) but not when it was abstract (Experiment 1). Taken together, these findings support the heuristic view of the effect of visibility on gesture production. That is, speakers have a heuristic that when they can see the addressee's face, they produce more representational gestures. The results are not compatible with the perspective taking view, which claims that the speaker produces more gestures when the addressee can see the speaker's gestures. Furthermore, speakers adjusted gesture rates only when talking about spatial content, but not when talking about abstract content. This indicates that speakers are sensitive to how useful their gestures might be for communication; that is, speakers modulate gesture rates in response to the face-visibility cue only when they talk about contents (e.g., spatial contents) for which the addressee is likely to benefit from gestures. Taken together, the findings suggest that a cue-based heuristic mechanism is responsible for the effect of visibility on gesture where the addressee's face and spatial content jointly act as a cue for the speaker to produce more representational gestures.

The findings suggest that the addressee's face acts as a cue for speakers to produce more gestures for the addressee. The use of the addressee's face as a cue fits with speaker's focus on the addressee's face in prior research. Addressees use facial expressions to provide feedback to the speaker, which can impact the speaker's communication (Bavelas et al., 2000; Bavelas & Gerwing, 2011) and speakers are sensitive to the addressee's gaze in conversation (C. Goodwin, 1986; M. H. Goodwin & Goodwin, 1986). As speakers are paying attention to the face to coordinate verbal communication any way, speakers can monitor the addressee's face for cues without any additional cognitive cost. Furthermore, speakers can check if the addressee is paying visual attention to the speaker, a pre-requisite for visual communication. Finally, in most everyday situations, when the addressee's face can be seen, the addressee can see the speaker's gestures. Visibility of the addressee's face is an ideal cue; it is low cost, it checks a key pre-requisite for visual communication, and it works in most cases.

The cue-based heuristic view can explain the previous findings of the effect of visibility on gesture. Previous research into the effect of visibility on gesture production has always confounded gesture visibility with facial visibility, (e.g. Alibali et al., 2001; Mol et al., 2009). When the addressee could see the speaker's gestures, the speaker could also see the addressee's face. According to the results of the current study, facial visibility cues the speaker to produce more gestures. Additionally, previous research with speakers discussing abstract topics did not produce more gestures when the addressee could see the speaker (e.g. Bavelas et al., 1992; Rimé, 1982) which can be explained by speakers not finding gestures as useful for communicating abstract topics. Therefore, the heuristic based on facial visibility and spatial content parsimoniously explain the results from the current study and previous studies on the effect of visibility on gesture.

The clear empirical evidence for cue-based audience design in the current study provides additional support for cue-based audience designs in other aspects of communication. Speakers can attenuate both speech and gesture when retelling a story to the same addressee a second time by using the identity of the addressee as a cue (Galati & Brennan, 2010, 2014). Speakers could orient gesture production towards the addressee by using the location of the addressee as a cue (Özyürek, 2002). Speakers could use the identity of the addressee as a cue to trigger infant directed speech (Snow, 1972). The Lombard effect can also be explained by the cue-based heuristic view. Speakers could use the presence of the addressee as a cue increase the modulations made in response to background noise. As long as the cues necessary to trigger a behaviour are present, it is plausible that an audience design behaviour can be explained by a cue-based heuristic mechanism. The previous literature argued for such cue-based heuristics purely on the theoretical basis that perspective taking would be too computationally costly, and thus computationally simpler heuristics must be at play. The current study for the first time provided direct empirical evidence for a cue-based heuristic for gesture production, which makes it more plausible that cue-based heuristics are also at play in other aspects of communication.

We speculate that eye contact may be one of the factors contributing to a cue that triggers speakers to produce more gestures for the addressee. In the shoulder height screen condition in Experiment 2, the speaker and the addressee could see each other's faces. In such situation, people tend to fixate on the addressee's eyes most of the time (Gullberg & Holmqvist, 2006). Previous research suggests eye contact has an important role in the effect of visibility on gesture. Speakers produce more gestures when they were able to make eye contact with the addressee than when the speaker and addressee could see each other's faces but not make eye contact (Mol et al., 2011). Mol et al. (2011) compared the rate at which speakers produced gestures when talking over webcams or using an “eye-catcher device” so that when the speaker and addressee are looking at each other on computer screens, it appears to the speaker and the addressee as if they are making eye contact. Speakers gestured more in the eye-catcher condition than the webcam condition. This suggests that eye contact contributes to a trigger for the heuristic. More broadly, we speculate that the face visibility cue concerns factors associated with the addressee's visual attention on the speaker.

The current study appears to contradict prior findings suggesting speakers are sensitive to whether the speaker’s gestures can be seen (Mol et al., 2011). Over a webcam, speakers produced more gestures when the addressee could see the speaker. Speakers did not produce more gestures when the speaker could see the addressee. The finding appears to suggest that speakers are sensitive to whether their own gestures are useful for communication, and do not respond to the addressee’s face as a cue. However, the findings of the previous study can be explained by the use of a web cam and the experimental instructions. The addressee’s face may not have acted as a cue because eye contact was not possible. When eye contact could be made, speakers produced far more gestures when the speaker knew their gestures could be seen by the addressee (Mol et al., 2011). Furthermore, speakers may have altered their gesture production after the experimenter highlighted whether the addressee could see the speaker or not by displaying the addressee’s point of view. It is plausible that showing the addressee’s perspective reduced the cognitive cost of perspective taking. Therefore, speakers may have produced more gestures in response to being shown by the experimenter that they were visible, or supressed gesture production when shown that they were not visible to the addressee.

The current study provides no evidence that the addressee’s visual feedback contributes to the effect of visibility on gesture production. Visual feedback from the addressee’s face and gesture visibility have been confounded in previous studies (see Bavelas and Healing, 2013, for a review). In both Experiment 1 and 2 of the current study, the addressee was judged to be more visually responsive in the addressee responsive condition than in the addressee non-responsive condition. Despite the visual responsiveness difference, speakers did not produce more gestures in the responsive condition than the non-responsive condition. Furthermore, the behaviour of the addressee was manipulated in the same way in both Experiment 1 and Experiment 2. If the addressee’s responsiveness is responsible for the effect of visibility on gesture, then both Experiment 1 and 2 should have found an effect of visibility on gesture.

The effect of visibility on gesture was dependent on the type of information the speaker was conveying. That is, the effect was found when the speaker was trying to convey dynamic spatial information about a cartoon, but not when the speaker was conveying abstract information about social situations. This suggests that speakers only designed their gestures for the addressee when gestures were especially suited to communicating the speaker's message (Hostetter, 2011). Gestures are particularly suited to conveying spatial information rather than abstract information because gestures can efficiently convey action, shape, relative location of objects and the movement of objects (Hostetter, 2011). Speakers produce gestures more when describing spatial information than non-spatial information (Alibali, 2005; Rauscher et al., 1996) and addressees benefit from seeing gestures conveying spatial and motoric gestures, but not from gestures conveying abstract information (Hostetter, 2011). These findings make it plausible that speakers are sensitive to how useful gesture is for communication and adapt gesture production accordingly. Considering our findings, we suggest that the heuristic mechanism triggers the speaker to produce additional gestures when discussing topics where gestures are especially suited to conveying information.

The effect of visibility on gesture only occurring when speakers are discussing dynamic spatial topics, but not abstract topics is consistent with previous studies. Previous studies that have found an effect of visibility had speakers describe cartoons (e.g. Alibali et al., 2001). In contrast, studies that have not found an effect of visibility on gesture have used stimuli conveying abstract information. For example, speakers discussed their favourite movie (Rimé, 1982) or told a close call story (Bavelas et al., 1992). Whether studies using spatial stimuli found an effect of visibility on gesture depends on the type of information being conveyed. Studies using dynamic spatial stimuli such as where participants gave directions elicited an effect (Cohen, 1977), but studies using static spatial stimuli where participants described pictures (e.g. pictures of a dress) did not (Bavelas et al., 2008; Holler et al., 2011). No study using abstract stimuli has found an effect of visibility on gesture (e.g. Bavelas et al., 1992; Rimé, 1982) and no study using dynamic spatial stimuli has failed to find an effect of visibility on gesture (e.g. Alibali et al., 2001). The current study furthers the idea that the effect of visibility on gesture is only found when speakers are describing stimuli which convey dynamic spatial information.

The current study is the first to provide evidence that speakers consider gestures to be more valuable for communication when conveying dynamic spatial information than when conveying abstract information. Speakers adapted their communication in response to the addressee's face when in the experiment where participants conveyed dynamic spatial information, but not in the experiment where participant's conveyed abstract information. This finding may also partially explain why speakers produce more gestures when conveying dynamic spatial information rather than abstract information (Feyereisen & Havard, 1999). This increase in gesture rate for dynamic spatial information has been attributed to the increase in motor imagery elicited when discussing dynamic spatial stimuli (e.g. Hostetter & Alibali, 2008). As far as we know, there is no previous evidence to suggest that speakers produce more gestures when conveying dynamic spatial information *for communicative reasons*.

Cue-based heuristics can be incorporated into a mechanistic framework for audience design (Ferreira, 2019). While Ferreira's framework is concerned with language audience design, it is possible to incorporate non-verbal behaviours into the same model. One source of audience design in the model comes from pre-learned automatic encoding strategies. Speakers can apply learned behaviours to create rules in a feedforward language production process. These rules establish the connection between a cue and a specific behaviour to form a heuristic mechanism. We suggest that the cue-based heuristic demonstrated in the current study fits into the mechanistic framework for audience design as a pre-learned automatic encoding strategy.

The perspective taking and cue-based heuristic mechanisms are not mutually exclusive. Speakers likely access a model of the addressee when the perspective of the addressee becomes important for completing a task. For instance, speakers are likely to take the addressee's perspective to determine if an adjective is necessary to distinguish a target item (e.g. a circle) from a distractor item (e.g. a larger circle) (Horton & Keysar, 1996). Speakers are more likely to provide a clarifying adjective when the distractor is in common ground than privileged ground. Under time pressure, speakers cease to adapt communication depending on the common-ground status of distractor items, suggesting speakers must take time to perspective take. We suggest that speakers will take the addressee's perspective when sufficiently motivated and have time to do so.

We did not replicate previous findings that speakers produce more gestures when the addressee provides responses to the speaker (Bavelas et al., 2008; Beattie & Aboudan, 1994). That is, speakers' gesture production was unaffected by the addressee's responsiveness behaviour. A stronger responsiveness manipulation may be required for an effect of addressee responsiveness. Previous responsiveness manipulations compared a silent addressee or an empty room/tape recorder to a conversational addressee. The current study compared a silent addressee to an addressee who provided controlled responses rather than freely engaging in conversation. The manipulation may not have been strong enough to elicit an effect.

The findings of the current study may have been influenced by differences in the addressee's behaviour between the different conditions. It is possible that the addressee may have used more of one type of verbal response than another. Furthermore, it is possible that the addressee may have used different visual responses which affect gesture production differently. Such differences will not have been captured by our control measure of addressee responsiveness rating. However, any such differences would have produced much smaller effects than the addressee responsiveness manipulation. If small variations in addressee responsiveness had a significant effect speaker's gesture production, then larger variations in addressee responsiveness should have had an effect on addressee responsiveness. As no effect of addressee responsiveness was found, it suggests that speaker's gesture production was not subject to variation in the addressee's responsiveness.

Further research using the current study's data set could examine the content of the gestures to establish if the information expressed in gesture changes in response to cues from the addressee. It's plausible that the content of gesture changes when speakers are communicating information to the addressee, especially as more informative gestures benefit communication (Krason et al., 2021). It is possible that while speakers gestured more when the addressee's face was visible to the speaker, the gestures may have only been more informative to the addressee when the speaker's gestures were visible. Further research could establish if the speaker's gestures are more informative when the gestures are visible to the addressee, or if the addressee's face is visible to act as cue to the speaker.

3.6. Conclusion

To summarise, the current study unequivocally showed that audience design behaviours can be triggered by cues from the addressee. Speakers produce more representational gestures in response to seeing the addressee's face but not in response to the addressee being able to see the speaker's gestures. We suggest that the face acts as a cue for the speaker to produce additional gestures to benefit communication. This effect of visibility on gesture was observed when speakers were describing dynamic spatial stimuli but was not observed when speakers discussed abstract stimuli. This finding suggests that speakers only utilise the addressee's face as a cue when gesture is particularly useful for communication such as when describing dynamic spatial stimuli. Taken together, speakers can use cue-based heuristics to produce audience design behaviours.

Financial Disclosure

This work was supported by the Midlands Graduate School ESRC DTP [grant number: ES/P000711/1]

Acknowledgements

We would like to thank Stéphen Theron-Grimaldi, Daniella Waluube, Sijjal Riaz, and Levent Emir Ozder for providing reliability coding and the Midlands Graduate School ESRC DTP for providing studentship funding.

Chapter 4

Gesture in Context: Investigating the Factors and Mechanisms that Affect Co-Speech Gesture Production

Abstract

Speakers naturally produce co-speech gestures to help communicate their message (Kendon, 1994). Speakers can produce gestures differently depending on the context. The current study investigated how two aspects of interactional context affected how frequently and prominently people produced gestures. The first aspect is to what extent the speaker and the addressee can see each other. Speakers produce more gestures overall when speaking to a visible addressee rather than an non-visible addressee (e.g. Alibali et al., 2001). The second aspect is whether or not the addressee provides feedback to the speaker. Speakers produce more co-speech gestures when talking to responsive addressees than non-responsive addressees (e.g. Beattie & Aboudan, 1994). The current study investigated whether the two aspects of the Interactional context have been confounded in previous research (see Bavelas & Healing, 2013, for a review). The current study also investigated whether speakers utilise cue-based heuristics or take the perspective of the addressee to design gestures to communicate more effectively with the addressee. We found that speakers can both respond to cues from the addressee using heuristics and take the perspective of the addressee to communicate more effectively with the addressee. Furthermore, we found that speakers produce more gestures when the addressee is visible, but speakers did not produce more gestures when the addressee was responsive. Furthermore, we found no evidence to suggest that the effect of visibility was due to the confounding of visibility and addressee responsiveness.

4.1 Introduction

Speakers produce hand and arm movements when they talk (McNeill, 1992). Co-speech gestures are spontaneously produced during speech and are typically co-ordinated with the words in timing and meaning (McNeill, 1992). Co-speech gestures are thought to be produced for two reasons. The first reason is to benefit the speaker producing the gestures (Kita et al., 2017), and the second is to help communicate information to an addressee (Kendon, 2004). Gestures can convey information relevant to speech to help an addressee understand the speaker's message (Hostetter, 2011; Riseborough, 1981). Gestures can convey additional information not expressed in speech, providing new information to the addressee (Broaders & Goldin-Meadow, 2010). Gestures can provide a supplementary source of information if the speaker does not comprehend the message being conveyed by speech (Sueyoshi & Hardison, 2005). Furthermore, gestures are especially effective at conveying spatial information (Alibali, 2005).

Speakers can produce different types of gestures. Representational gestures depict an action, movement, object, shape, location, or metaphorically depict abstract concepts. For instance, a speaker could move their hand from one side to another while talking about a character crossing the street. The lateral movement represents the path made by the character. Interactive gestures support the interaction with the addressee and directly reference the addressee (Bavelas et al., 1992, 2008). For instance, a speaker could offer the addressee a turn in the conversation by offering out a hand towards toward the addressee.

Speakers produce gestures differently depending on the context. For example, when conveying information on a topic the addressee does not know much about, speakers produce more gestures (Jacobs & Garnham, 2007), and also make the gestures larger (Holler & Stevens, 2007). Speakers also produce larger gestures when motivated to communicate clearly and effectively (Hostetter et al., 2011). Furthermore, speakers produce more gestures when the addressee can see the gestures than when the gestures are hidden (Alibali et al., 2001).

The current study investigated how two aspects of interactional context affected how frequently and prominently people produced gestures. The first aspect is to what extent the speaker and the addressee can see each other. The second aspect is whether or not the addressee provides feedback to the speaker.

4.1.1 The Effect of Visibility on Gesture

One part of the conversational context that can affect gesture production is whether the addressee can see the speaker's gestures. Speakers produce more gestures overall when speaking to a visible addressee rather than an non-visible addressee (Emmorey & Casey, 2001; Hostetter & Potthoff, 2012; Krauss et al., 1995; Mol et al., 2009a, 2009b). Furthermore, speakers produce more representational or illustrator gestures (Alibali et al., 2001; Cohen, 1977; Cohen & Harrison, 1973) and more interactive gestures (Bavelas et al., 1992, 2008) when speaking to a visible addressee than an non-visible addressee. The difference in gesture rate has been attributed to the communicative function of gestures. When the addressee cannot see the speaker's gestures, speakers are not motivated to communicate using gesture. When the addressee can see the speaker's gestures, then speakers are motivated to communicate using gesture.

Speakers also alter the form and location of their gestures when the addressee can see the speaker's gestures. Speakers produce larger representational gestures when the addressee can see the speaker (Bavelas et al., 2008). Speakers also placed gestures higher up in space when the addressee was visible (Holler et al., 2011). Speakers may produce larger and higher up gestures when the speaker wants to communicate information using gesture. The larger gestures may be clearer and help the speaker convey information, while higher up gestures may be able to catch the addressee's attention.

4.1.2 *Mechanisms to explain the effect of visibility on gesture*

Speakers may produce more gestures when talking to a visible addressee because speakers can take the perspective of the addressee. To take the perspective of the addressee, speakers build and maintain a model of the addressee and accesses this model when necessary (Clark & Marshall, 1981; Clark & Murphy, 1982). Speakers constantly update this model over time as new information about the addressee becomes available (Krauss & Fussell, 2004). According to a perspective taking view, the speaker takes the addressee's perspective to infer if the addressee can see the speaker or not. When the speaker infers that the addressee can see the speaker's gestures, then the speaker produces more gestures.

Speakers may produce more gestures when talking to a visible addressee because speakers can respond heuristically to cues from the addressee. Maintaining a model of the addressee's knowledge is cognitively demanding (Rossnagel, 2000) and time consuming (Epley et al., 2004; Horton & Keysar, 1996). Speakers can instead take advantage of cues available in their interaction with the addressee (Shintel & Keysar, 2009). Such cues can trigger speakers to change their behaviour in pre-determined ways. According to a cue-based heuristic view, speakers produce more gestures in reaction to a cue related to the addressee. Speakers produce more representational gestures when the addressees face is visible, even when the speaker's gestures were not visible to the addressee (Barker & Kita, 2021), suggesting the addressee's face can be a cue to trigger the speaker to produce more gestures.

4.1.3 *Motivating speakers to take the perspective of the addressee*

No previous study has provided unequivocal evidence for speakers taking the perspective of the addressee rather than automatically producing behaviours in response to cues. Previous studies have not separated mutual gesture visibility from mutual facial visibility. Without separating gesture visibility and facial visibility, we cannot attribute the effect of visibility on gesture to speaker's inferring their gestures can be seen and being motivated to communicate using gesture. In the only direct comparison of the perspective taking and cue-based heuristic, no support for the perspective taking explanation was found (Barker & Kita, 2021). Speakers did not produce more representational gestures when the speaker's gestures were visible to the addressee, but the addressee's face was not visible to the speaker. Speakers did not produce enough interactive gestures when talking to the experimenter to conduct an analysis. Furthermore, speakers did not produce gestures at different heights when the addressee was visible or not, suggesting speakers did not infer gestures could not be seen and decide to raise the height of gestures above the shoulder height screen.

It is plausible that a speaker who is motivated to communicate effectively is more likely to take the perspective of the addressee than an unmotivated speaker. If speakers are more motivated to communicate effectively, then speakers may be more willing to allocate cognitive resources to taking the perspective of the addressee. It is plausible that speakers may rely on cue-based heuristics for audience design behaviours until speakers are sufficiently motivated to take the perspective of the addressee. Participants are more likely to accurately take the perspective of an addressee when there is a strong motivation to accurately guess how the addressee will interpret a phrase (Epley et al., 2004). Thus, it is plausible that speakers who are motivated to take the perspective of the addressee will attribute more cognitive resources to perspective taking. Speakers may have needed to have been sufficiently motivated to take the perspective of the addressee as perspective taking can be cognitively demanding (Rossnagel, 2000). The suggestion that speakers only take the perspective of the addressee when motivated to do so fits with predictions of the interactive-alignment model (Pickering & Garrod, 2004). According to Pickering and Garrod, (2004), perpsective taking is a costly strategy which is not typically needed as more simple stragetgies work the majority of the time.

The current study may provide evidence for a perspective taking explanation of the effect of visibility on gesture. In the previous study (Barker & Kita, 2021), speakers described stimuli to the experimenter. It is possible that speakers assumed the knowledge state of the experimenter. Speakers may have believed the experimenter was already knowledgeable about the cartoons, and that the experimenter was not under pressure to correctly recall the story afterwards. Therefore, speakers may have been less motivated to communicate information effectively to the experimenter. In the current study, speakers describe stimuli to naïve participants. It is plausible that speakers will believe it will be important to communicate the information to the fellow participant. Speakers may believe the student addressee was ignorant about the cartoons, and that the addressee will be under pressure to correctly recall the story afterwards. If speakers believed it was important to clearly convey the information to the addressee, then speakers may have been more motivated to take the perspective of the addressee to ensure effective communication.

It is also plausible that speakers will be more motivated to take the perspective of the addressee when perspective taking is less cognitively demanding. Speakers may be more likely to take the addressee when the addressee is known to themselves because the cost of perspective taking is reduced. If speakers already share a great deal of common ground, then taking the perspective of peers may not be as costly as taking the perspective of strangers. Members of the same community may be more likely to correctly interpret ambiguous utterances than those of distinct communities (Clark & Marshall, 1981). Speakers are more likely to interpret ambiguous statements correctly in line with community expectations when speaking to a friend than a stranger (Gerrig & Littman, 1990). Furthermore, participants are more accurate at identifying abstract stimuli when reading a description written by a friend rather than a stranger (Fussell & Krauss, 1989).

Taken together, it is plausible that when describing stimuli to fellow participants, speakers will take the perspective of the addressee in the current study. Previously, perspective taking and cue based heuristic explanations have only been compared with participant and experimenters interacting (Barker & Kita, 2021). It is plausible that speakers are more able or willing to take the perspective of peers than an experimenter as participants may feel more comfortable speaking to a peer and believe they have more in common.

4.1.4 The effect of addressee responsiveness on gesture

A further part of the conversational context that can affect gesture production is how the addressee provided feedback to the speaker. The behaviour of the addressee can affect speakers' gesture production. Speakers produce more co-speech gestures when talking to responsive addressees than non-responsive addressees (Beattie & Aboudan, 1994). When addressees provide feedback to the speaker, then speakers produce more gestures in response to the addressee to confirm the message is understood correctly (Bavelas et al., 2011). Speakers also produce fewer gestures when communicating with a less attentive addressee (Jacobs & Garnham, 2007).

Speakers may produce more gestures when talking to an addressee who provides feedback because speakers can react to the feedback from the addressee using gesture. Speakers and addressees can provide information back and forth between them to ensure the communication is effective (Schober & Clark, 1989). The speaker communicates a piece of information, the addressee can signal their understanding and the speaker can confirm the understanding is correct. Addressees can provide verbal responses such as "mm-hmm" and "yeah" as part of this back channel of communication between the speaker and listener (Yngve, 1970). Addressees can also provide non-verbal responses (Bavelas et al., 2011). Gaze is used to create a back-channel response through which messages can be sent (Bavelas et al., 2002). Furthermore, addressees can use facial expressions to respond to the speaker. These include smiles (Brunner, 1979) and mimicking the speaker's facial expressions (Bavelas, 2007; Bavelas et al., 2000). Speakers can then use gesture to respond to the addressee's feedback during conversation.

4.1.5 The addressee responsiveness explanation for the effect of visibility on gesture

Speakers may produce more gestures when talking to a visible addressee because the addressee can provide non-verbal feedback to the speaker (see Bavelas & Healing, 2013, for a review). When speakers and addressees cannot see each other, then both the speaker's gestures and the addressee's visual responses to the speaker are blocked. Thus, the speaker's gestures being visible or not is confounded with whether that addressee's responses are visible or not. When the addressee's visual responses are hidden from speakers, then speakers may produce fewer gestures. Thus, it is possible that speaker produce fewer gestures when the addressee is hidden behind a screen because the addressee cannot produce visual responses to the speaker, rather than because the speaker cannot use gestures to communicate information to the addressee.

Studies may not find an effect of visibility on representational gesture rate when the addressee is able to provide verbal responses. Several studies have failed to find the effect of visibility on gesture (Bavelas et al., 1992, 2008, 2014b; de Ruiter et al., 2012; Holler et al., 2011; Pine et al., 2010; Rimé, 1982). The difference in findings between studies that find an effect of visibility and those that do not find an effect can be explained by methodological differences between the studies. In studies where the effect of visibility on gesture is found, the verbal behaviour of the addressees is controlled and standardised (quasi-dialogues designs). In quasi-dialogue designs, the addressees are typically researchers or confederates who give limited verbal responses. When the speaker and addressee are prevented from seeing each other, then the addressee's visual responses to the speaker cannot be seen. Thus, when the speaker's gestures are not visible, the addressees are also not responsive. In studies where the effect of visibility on gesture is not found, the verbal behaviour of the addressees is not controlled (free-dialogue designs). In free-dialogue studies, the addressees are allowed and sometimes encouraged to discuss the stimuli with the speaker. In free-dialogue studies when the addressee is not visible to the speaker, the addressee may naturally compensate for not being able to provide visual by providing additional verbal responses. If the addressee keeps the level of feedback to the speaker constant when gesture is both visible and non-visible, then there should be no effect of visibility on the number of representational gestures produced (e.g. Bavelas et al., 2008).

It is plausible that responses from the addressee could explain the effect of visibility on representational gesture. If feedback from the addressee is being systematically confounded with manipulations of gesture visibility, then it is not clear whether visibility of gesture or addressee responsiveness causes the effect of visibility on representational gestures.

There is little empirical evidence to support the claim that addressee responsiveness is responsible for the effect of visibility on representational gestures. If addressee responses can explain the effect of visibility on representational gesture production, then speakers should only produce more gestures to a visible addressee than a non-visible addressee when the addressee's responses are limited and controlled. No previous study has compared a conversational addressee to an addressee whose behaviour is controlled while also manipulating visibility. Previous studies have compared a visible addressee to both a non-visible addressee and a tape recorder (Bavelas et al., 2008), but this design cannot establish an effect of visibility with a controlled addressee. To support the addressee responses explanation (Bavelas & Healing, 2013), the effect of visibility must be compared between having a controlled non-responsive addressee and a conversational responsive addressee.

4.1.6 The current study

The current study investigated the effect of visibility on speaker's gesture production. We hypothesise that a heuristic mechanism would utilise the addressee's face as a cue for speakers to use representational gesture for communication. More specifically, speakers should produce more representational gestures when there is no visual barrier between the speaker and the addressee because they can see the addressee's face, but not because the addressee can see the speaker's gestures. We hypothesise that speakers will take the perspective of the addressee when producing interactive gestures. More specifically, speakers will infer whether the addressee can see the speaker's gestures and produce more interactive gestures when the gestures are visible. We further hypothesise that speakers will only adjust the hight and saliency of their gestures when taking the perspective of the addressee. More specifically, speakers will produce gestures higher up and more saliently when the gestures can be seen by the addressee.

To distinguish between when the speakers' gestures were visible and when the addressee's face was visible, we introduced a condition in which a shoulder height screen was placed between the speaker and the addressee to prevent gestures being seen by the addressee while allowing the speaker and addressee to see each other's faces. In this condition, the addressee's face was still visible to the addressee, but speakers' gestures were not visible to the addressee.

We investigated the effect of visibility in three conditions. In a first condition the speaker communicated with the addressee face to face with no screen, so the speaker and addressee could see each other without any restrictions. In a second condition the speaker communicated with the addressee with a shoulder height screen between them, so the addressee could not see the speaker's gestures, but the speaker and addressee could see each other's faces. In a third condition the speaker communicated with the addressee with a full height (above head) screen between them, the speaker and addressee could not each other at all.

We predict that speakers should produce more representational gestures in the no screen and shoulder height screen conditions than in the full screen condition. Speakers should also produce more interactive gestures in the no screen condition than in the shoulder height and full screen conditions. Furthermore, speakers should produce representational gestures higher up and more saliently in the no screen condition than in the shoulder height and full screen conditions

The current study also investigated the effect of addressee responsiveness on speaker's gesture production. We hypothesise that speakers will produce more representational and interactive gestures when speaking to an addressee who responds and provides feedback. We manipulated the addressee's responses in two conditions. Half of the addressees were asked to be conversational and provide responses to the speaker. The other half of the addressees were asked to be silent. We predict that speakers will produce more representational and interactive gestures in the conversational condition than in the silent condition.

The current study also investigated if the effect of responsiveness can explain the effect of visibility on gesture. Speakers should produce more gestures when talking to a conversational addressee than a silent addressee. Therefore, we predict that speakers should produce more representational and interactive gestures in the conversational condition than in the silent condition.

If Bavelas and Healing's (2013) explanation for the effect of visibility on gesture production is correct, then speakers should produce more representational gestures when the addressee is responsive. We should see an interaction between the addressee's responsive behaviour and the mutual visibility between the speaker and addressee. Speakers should produce more representational gestures in the no screen and shoulder height screen conditions than in the full screen condition, but only in the silent condition, not in the conversational condition.

4.2 Methods

4.2.1 Participants

36 participants took part as speakers and a further 36 participants took part as addressees. Participants were first year Psychology undergraduate students from the University of Warwick who received course credit. Both speakers and addressees reported English as a first language. Two pairs of participants were excluded from the analysis. One pair were excluded due to the addressee gesturing. Another pair were excluded due to the speaker not reporting English as their first language. The data from the remaining 34 pairs of participants were analysed. The number of participants recruited was based on previous studies investigating the effect of visibility on gesture (Barker & Kita, 2021; Bavelas et al., 2008) in which 30 participants were analysed.

4.2.2 Design

The experiment used a 2x3 mixed factor design with one within-subject independent variable, Visibility (no screen, shoulder height screen, full screen), and one between-subject independent variable, Addressee responsiveness (silent, conversational). Four dependant variables were the rate of representational gestures per 100 words, the rate of interactive gestures per 100 words, the average representational gesture height, and the average representational gesture saliency.

4.2.3 Materials

Participant's speech and gesture were elicited by video clips of six episodes from a Tweety and Sylvester film entitled "Canary Row" (which was also used in McNeill, 1992).

Visibility was manipulated by adjusting a screen fixed between two height-adjustable tripods. See Figure 1.

Two Canon Legria R66 video cameras were used to record during the experiment. The recordings of the experimenter were used as stimuli for participants to provide the addressee's visual responsiveness rating.

Figure 1

The experimental set up. The speaker sat on the chair on the right bottom, and the addressee sat on the other side of the screen. The laptop is situated next to the participant. One camera recorded the participant, and a second camera recorded the experimenter. The top of the screen was adjusted for different conditions by changing the height of the tripod, while the bottom of the screen was kept 10cm above the floor.

4.2.4 Procedure

The experimenter informed the participants that they would be taking part in an experiment about describing cartoons. The experimenter did not inform the participants that the study was investigating gesture. Participants were assigned to the role of speaker and addressee based on a coin flip. The experimenter assigned the participant pairs to either the silent or conversational condition alternatively. The first pair were assigned to the silent condition and the second pair were assigned to the conversational condition, etc. The participants were seated opposite from one another in chairs. The stimuli were displayed to the side of the speaker on a laptop.

The participants took part in a practice cartoon description with the experimenter present and three cartoon descriptions with the experimenter absent. The participant assigned to the role of the speaker were shown how to start each video clip and move on to the next on the laptop. The laptop was placed on a table to the side of the speaker. The speaker used headphones to listen to the audio of the stimuli and was instructed to watch a video clip twice. The speaker then placed the headphones back on the table and turned to face the addressee and described the cartoon in as much detail as possible. Once the speaker had finished describing the cartoon, the speaker fetched the experimenter from outside the room. This repeated for all three conditions.

In the *no screen* condition, the screen was set aside so that the participant and experimenter could both see each other with no barrier. In the *shoulder height screen* condition, the screen was placed between the experimenter and participant, and the top of the screen was raised to the shoulder height so that they could only see each other's chins and above. In the *full screen* condition, the screen was placed between the experimenter and participant, and the top of the screen was raised to the height such that they could not each other at all.

In the *silent* condition, the experimenter instructed the participant in the role of the addressee to remain silent throughout. If the participant spoke during the practice cartoon description, they were reminded that they should remain silent. No participants in the silent condition spoke during the actual task. In the *conversational* condition, the addressee was asked to try and have a conversation with speaker about the cartoon, including asking questions and making comments, to try and get as much detail and as complete an understanding of the video as possible. All the participants in the conversational condition spoke to some degree. We did not specify a minimum required amount of conversation to keep the interactions natural and the instructions similar to those in previous studies (e.g. Bavelas et al., 2008)

The speakers watched each cartoon and described it to the addressee. Participants did this for all three cartoons. We counterbalanced condition order so that the three visibility conditions were presented equally often in all possible (six) orders for both silent pairs and conversational participant pairs

Participants were then asked to fill in a short questionnaire to obtain handedness information and biographical information as well as a full video consent form, Participants were then fully debriefed.

4.2.5 Gesture Coding

Participants' gestures were coded using a modified version of the coding manual used by Chu et al. (2014). Participants' gestures were coded during the stroke phase of the preparation-stroke-hold-retract breakdown of gestures (Kita et al., 1997). If there was a clear preparation or hold between two strokes, they were coded as separate gestures. Gestures were categorised based of their form and function for analysis. The two key gesture types were representational gestures and interactive gestures. Representational gesture and interactive gesture each consisted of several sub-categories of gesture. The coding criteria are in "Gesture Coding Manual" in the supplementary materials.

Representational gestures are made up of depictive, deictic and conduit gestures. Depictive gestures are movements which can depict the actions of an agent, the movement of an object or a property of either an actor or agent. This category subsumes "iconic", "metaphorical", and "deictic" gestures, as defined by McNeill (1992). Iconic gestures depict some physical action, object, or aspect of an object. Metaphorical gestures depict a metaphorical concept. Deictic gestures are pointing gestures which identify a location in the space. Some representational gestures are classified as representationally unclear if they have no clear meaning but still seem to be shaped to represent something. Conduit gestures (McNeill, 1992) are produced by the speaker moving an open palm towards the addressee to present an idea. The speaker should be certain about what they are saying. Conduit gestures have a clear forward movement of the hand toward the addressee.

Interactive gestures (Bavelas et al., 1992) are made up of listener indexing, conduit and palm revealing emblem gestures. Listener-Indexing gestures refer to the addressee. A gesture is coded as a listener-indexing gesture when the hand moving towards the addressee and accompany speech that references the addressee. For listener indexing gestures, previous research had stated that speaker would typically make eye contact with the addressee (Chu et al., 2014), but this criterion was dropped because in the full screen condition, the speaker and addressee could not make eye contact. Conduit gestures (as seen in the previous paragraph) were coded as interactive gestures as well as representational gestures. This is because conduit gestures directly index the addressee, as the idea being presented is being placed on the palm and moved toward the addressee. Palm revealing gestures included shrugs with the palms out and facing upwards to express empty handedness. They indicate uncertainty, or a lack of knowledge to share.

Other gesture classifications were coded, but not included in the calculation of dependant variables in the current study. These included beat gestures which are small movement up and down, typically in the lap and the hand shape is loose. There should be no clear meaning assignable to the gesture. Emblems are symbolic gestures where a hand movement has a conventional meaning. An example of this being a thumbs-up emblem. Metacognitive gestures are repetitive movement that indicate thinking. Generally, this involves repeated finger tapping or rotating the hand at the rest while pausing speech to search for a word or phrase. Unclear gestures are those that fail to fit into any of these categories.

Representational gestures were coded into three different heights. The first height category was “lower than waist”. The waist was defined as being part way between a person’s hips and the bottom of their ribcage. The second height category was “between waist and chin”. The third height category was “above chin”. For any given gesture, an imaginary line is drawn across the participant across their waist and another line is drawn across the base of the chin. The lines were drawn parallel to horizontal shelves in the background to account for the angle of the camera. The height of a gesture was determined by the highest point of the hand or figure during the gesture movement. The location of the highest point is compared to the waist and chin lines to determine the height coding. Lower than waist gestures were given a value of 1. Between waist and chin gestures were given a value of 2. Above chin gestures were given a value of 3.

Representational gestures were coded into four different categories of saliency. Gesture saliency was coded depending on which joints were involved in gesture. The first saliency category was “whole arm”. In this type of gesture, participants moved the upper arm at the shoulder joint. The second saliency category was “forearm”. In this type of gesture, participants moved the forearm at the elbow joint, and did not move the upper arm at the shoulder. The third saliency category was “hand”. In this type of gesture, participants moved the hand at the wrist, and did not move the forearm at the elbow, or the upper arm at the shoulder. The fourth saliency category was "fingers". In this type of gesture, participants moved their fingers at the knuckle joint, and did not move the hand at the wrist, the forearm at the elbow, or the upper arm at the shoulder. Finger gestures were given a value of 1, hand gestures were given a value of 2, forearm gestures were given a value of 3, and whole arm gestures were given a value of 4.

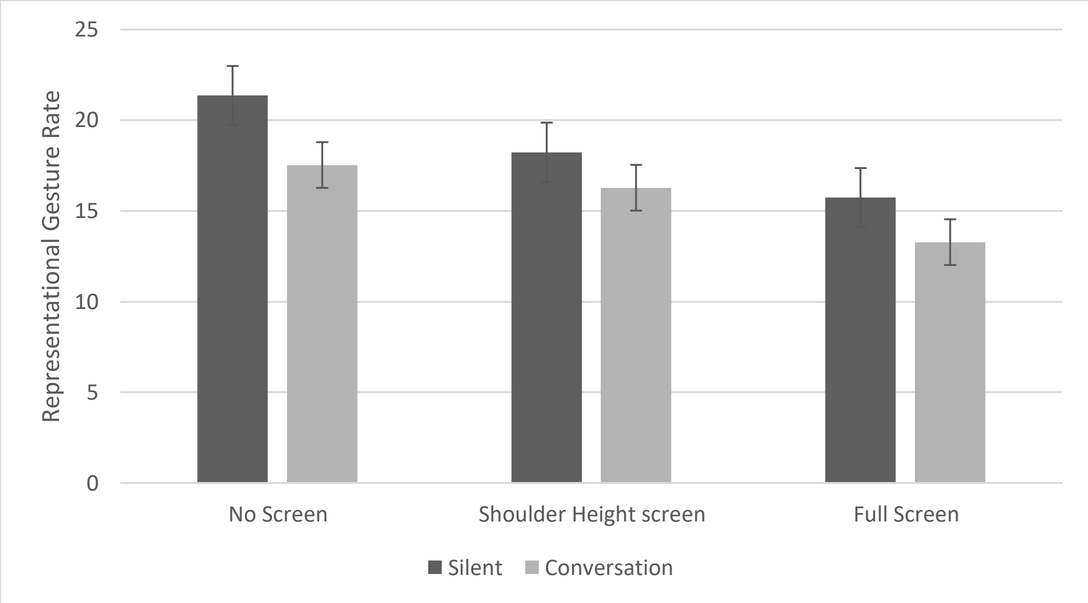
4.3 Results

The representational gesture rate was analysed in a 2x3 analysis of variance with visibility (no screen, shoulder height screen, full screen) as a within-participant independent variable and addressee responsiveness (silent, conversational) as a between-participant independent variable (See Figure 2 for descriptive statistics). There was a statistically significant main effect of visibility, (Greenhouse-Geisser corrected degrees of freedom) $F(1.66, 53.07) = 14.40$, $p < .001$, $\eta_p^2 = .310$, but no significant main effect of addressee responsiveness, $F(1, 32) = 3.74$ p = .062, $\eta_p^2 = .105$. There was no statistically significant interaction effect between visibility and addressee responsiveness, F(1.66, 53.07) = 0.55, p = .548, $\eta_p^2 = .017$.

To further explore the main effect of visibility, we ran posthoc pair-wise comparisons (as advised by Howell (2017) for a comparison of three means, we used Fisher's LSD test). Speakers produced representational gestures at a significantly higher rate in the no screen than the shoulder height screen condition $t(33) = 2.83$, $p = .008$, $d = 4.86$, 95% CI [0.13, 0.84] and the full screen condition $t(33) = 6.02$, $p < .001$, $d = 1.033$, 95% CI [0.61, 1.45]. Speakers also produced representational gestures at a significantly higher rate in the shoulder height screen condition than in the full screen condition $t(33) = 2.50$, $p = .018$, $d = 0.429$, 95% CI [0.07, 0.78].

Figure 2

Mean representational gesture rate (per 100 words) in the no screen, shoulder height screen, and full screen conditions when the addressee was conversational or silent. The error bars represent standard errors.

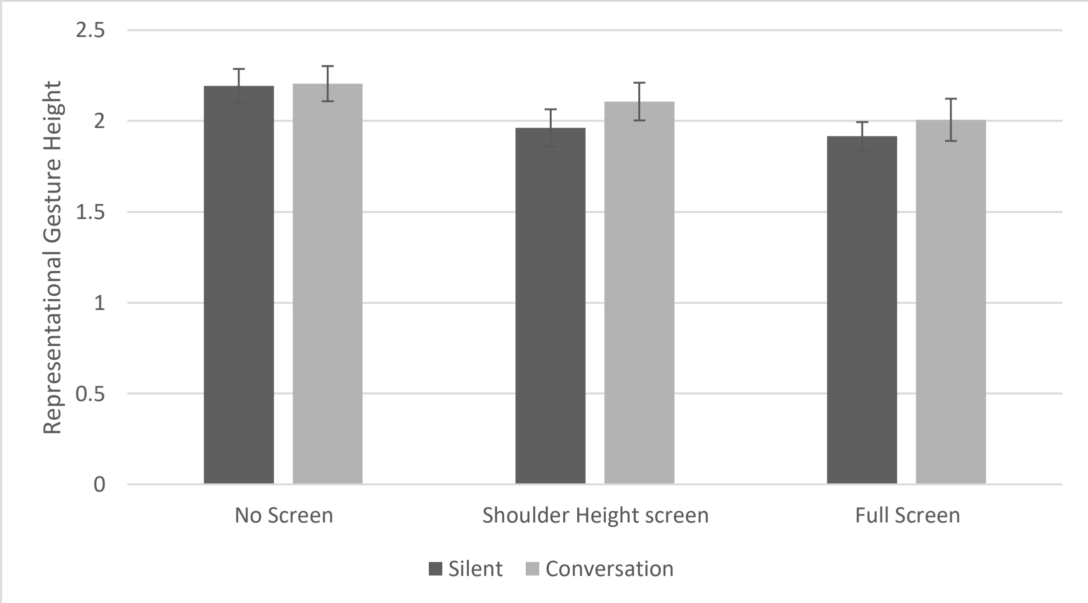


We calculated the mean height of representational gestures for each condition for each participant. The representational gesture height was analysed in a 2x3 analysis of variance with visibility (no screen, shoulder height screen, full screen) as a within-participant independent variable and addressee responsiveness (silent, conversational) as a between-participant independent variable (See Figure 3 for descriptive statistics). There was a statistically significant main effect of visibility, $F(2, 62) = 6$ (5.997), $p < .004$, $\eta_p^2 = .162$, but no significant main effect of addressee responsiveness, $F(1, 31) = 0.66$ p = .442, $\eta_p^2 = .021$. There was no statistically significant interaction effect between visibility and addressee responsiveness, F(2, 62) = 0.43, p = .650, $\eta_p^2 = .014$.

To further explore the main effect of visibility, we ran posthoc pair-wise comparisons (as advised by Howell (2017) for a comparison of three means, we used Fisher's LSD test). Speakers produced representational gestures significantly higher in the no screen than the shoulder height screen condition $t(32) = 2.93$, $p = .006$, $d = 0.510$, 95% CI [0.143, 0.870] and the full screen condition $t(32) = 3.38$, $p = .002$, $d = 0.589$, 95% CI [0.215, 0.955]. Speakers did not produce representational gestures significantly higher in the shoulder height screen condition than in the full screen condition $t(32) = 0.85$, $p = .404$, $d = 0.147$, 95% CI [-0.197, 0.489].

Figure 3

The mean representational gesture height in the no screen, shoulder height screen, and full screen conditions when the addressee was conversational or silent. The error bars represent standard errors.

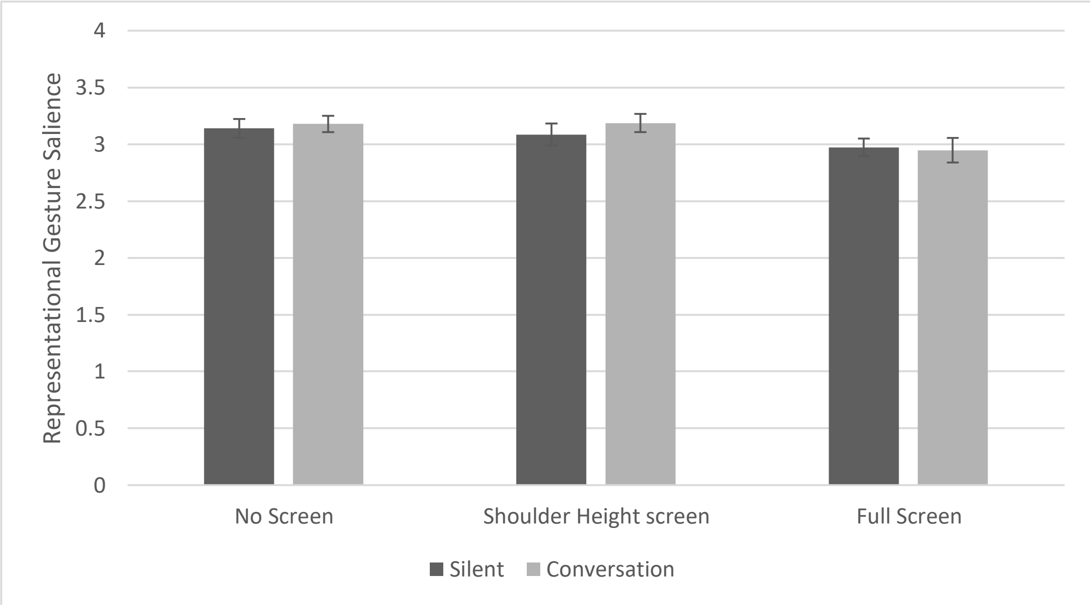


We calculated the mean salience of representational gestures for each condition for each participant. The representational gesture salience was analysed in a 2x3 analysis of variance with visibility (no screen, shoulder height screen, full screen) as a within-participant independent variable and addressee responsiveness (silent, conversational) as a between-participant independent variable (See Figure 4 for descriptive statistics). There was a statistically significant main effect of visibility (Greenhouse-Geisser corrected degrees of freedom), $F(1.46, 45.25) = 8.65$, $p < .002$, $\eta_p^2 = .218$, but no significant main effect of addressee responsiveness, $F(1, 31) = 0.129$ $p = .722$, $\eta_p^2 = .004$. There was no statistically significant interaction effect between visibility and addressee responsiveness, $F(1.46, 45.25) = 0.72$, $p = .451$, $\eta_p^2 = .023$.

To further explore the main effect of visibility, we ran posthoc pair-wise comparisons (as advised by Howell (2017) for a comparison of three means, we used Fisher's LSD test). Speakers produced representational gestures significantly more saliently in the no screen than the full screen condition $t(32) = 3.29$, $p = .002$, $d = 0.573$, 95% CI [0.201, 0.938]. Furthermore, speakers produced representational gestures significantly more saliently in the shoulder height screen condition than in the full screen condition $t(32) = 0.2.96$, $p = .006$, $d = 0.516$, 95% CI [0.148, 0.876]. Speakers did not produce representational gestures significantly more saliently in the no screen condition than in the shoulder height screen condition $t(32) = 0.75$, $p = .459$, $d = 0.130$, 95% CI [-0.213, 0.472].

Figure 4

The mean representational gesture salience in the no screen, shoulder height screen, and full screen conditions when the addressee was conversational or silent. The error bars represent standard errors.

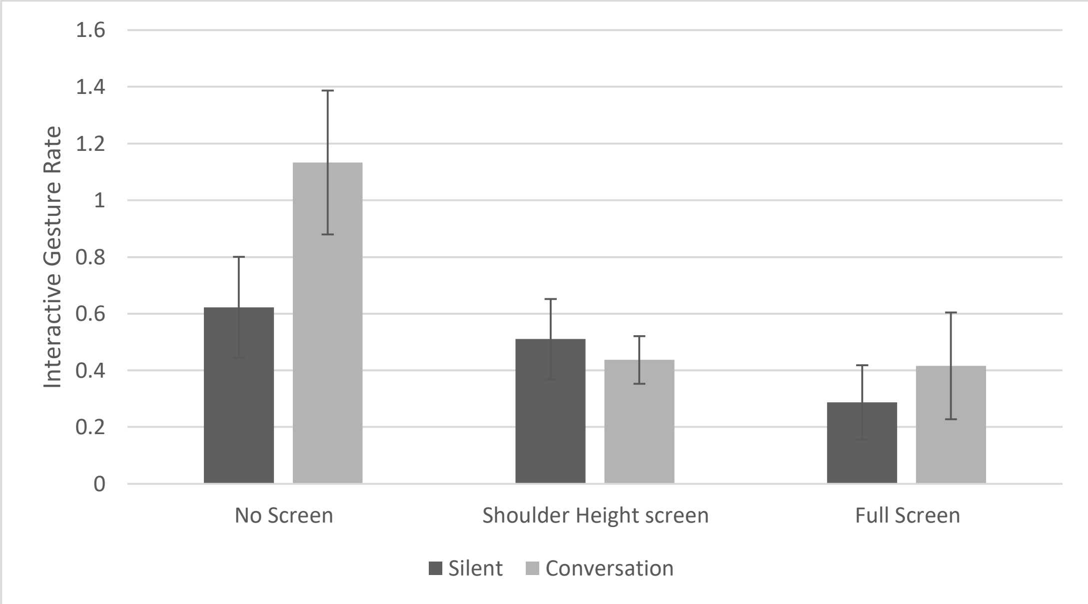


The interactive gesture rate was analysed in a 2x3 analysis of variance with visibility (no screen, shoulder height screen, full screen) as a within-participant independent variable and addressee responsiveness (silent, conversational) as a between-participant independent variable (See Figure 5 for descriptive statistics). There was a statistically significant main effect of visibility, $F(2, 64) = 6.91$, $p = .002$, $\eta_p^2 = .178$, but no significant main effect of addressee responsiveness, $F(1, 32) = 1.262$ $p = .270$, $\eta_p^2 = .038$. There was no statistically significant interaction effect between visibility and addressee responsiveness, $F(2, 64) = 2.01$, $p = .143$, $\eta_p^2 = .059$.

To further explore the main effect of visibility, we ran posthoc pair-wise comparisons (as advised by Howell (2017) for a comparison of three means, we used Fisher's LSD test). Speakers produced interactive gestures at a significantly higher rate in the no screen than the shoulder height screen condition $t(33) = 2.50$, $p = .018$, $d = 0.429$, 95% CI [0.074, 0.777] and the full screen condition $t(33) = 3.575$, $p = .001$, $d = 0.613$, 95% CI [0.242, 0.976]. Speakers did not produce interactive gestures at a significantly higher rate in the shoulder height screen condition than in the full screen condition $t(33) = 0.84$, $p = .405$ $d = 0.145$, 95% CI [-0.194, 0.481].

Figure 5

Mean interactive gesture rate (per 100 words) in the no screen, shoulder height screen, and full screen conditions when the addressee was conversational or silent. The error bars represent standard errors.



4.4 Discussion

There are five key findings. First, speakers produced more representational gestures in the no screen condition than the shoulder height or full screen conditions, and more representational gestures in the shoulder height screen condition than the full screen condition. This effect of visibility was found regardless of the addressee silent and conversational conditions. Second, speakers did not produce significantly more representational gestures in the conversational condition than in the silent condition. Third, speakers produced more interactive gestures in the no screen condition than in the shoulder height and full screen conditions. Fourth, speakers produced representational gestures higher up in the no screen condition than in the shoulder height and full screen conditions. Finally, speakers produced gestures which were more salient in the no screen condition and shoulder height than the full screen condition. These findings have important implications for the explanation of the effect of visibility on gesture, as we detail below.

The findings suggest that speakers can both take the perspective of the addressee and alter communicative behaviours in response to cues from the addressee. Speakers produced more representational and interactive gestures and placed their gestures higher up when the speaker's gestures were visible to the addressee. The findings suggest that speakers were able to take the perspective of the addressee and infer whether gestures would be useful for communication. However, speakers produced more representational gestures and made their gestures more salient when the addressee's face was visible than when the addressee's face was not visible. The findings suggest that speakers may rely on cue-based heuristics for some audience effects. Together, these findings indicate that speakers produce more gestures in response to cues from the addressee and can also take the perspective of the addressee to infer if gestures can be used to communicate.

The current study is the first to provide evidence that speakers can both respond to cues from the addressee using heuristics and take the perspective of the addressee to communicate more effectively with the addressee. Previous findings have been compatible with a perspective taking explanation (e.g. Alibali et al., 2001), but the findings could also be explained by a cue-based heuristic explanation. The only previous comparison of a perspective taking and heuristic based explanation found no evidence that speakers took the perspective of the addressee (Barker & Kita, 2021). The current study extends the findings of audience design mechanisms by suggesting that speakers can both respond to cues and take the perspective of the addressee to adjust the same communicative behaviour. In the current experiment, speakers may have produced more gestures in response to seeing the addressee's face and taken the perspective of the addressee. In the no screen and shoulder height screen condition, speakers would produce more gestures as the addressee's face was visible, but at times may have taken the addressee's perspective and infer whether gestures could be seen or not. In the no screen condition, the speaker would produce more gestures when perspective taking, but in the shoulder height screen condition the speakers would not produce more gestures when perspective taking. Hence, speakers gesture more in the no screen condition than in the shoulder height screen condition, but also more in the shoulder height screen condition than in the full screen condition.

The current study is the first to provide evidence for the effect of visibility on representational gesture rate in a free-dialogue setting. Previous studies that have unconstrainted participants acting as addressees have failed to find an effect of visibility on gesture rate (Bavelas et al., 1992, 2008; de Ruiter et al., 2012; Holler et al., 2011; Pine et al., 2010; Rimé, 1982). The findings suggest that during conversation, speakers produce gestures to benefit communication when the gestures are visible to the addressee. The current study extends previous findings that speakers produce gestures to communicate when gestures are visible to the addressee in quasi-dialogue designs (e.g. Alibali et al., 2001) to free-dialogue designs.

The findings provided no evidence to suggest that the effect of visibility was due to the confounding of visibility and addressee responsiveness, as suggested by Bavelas and Healing (2013). First, the effect of visibility was found in both the silent and conversational addressee responsiveness conditions. There is no evidence that the effect of visibility on representational gesture only occurs when addressees are not able to be responsive as speakers produced more gestures when the addressee was visible than not visible even when addressees were free to be responsive in the conversational condition. Second, there was no effect of the addressee's responsiveness on the speaker's representational gesture rate. There was no evidence to suggest that responses from the addressee cause the speaker to produce more gestures. Together, these findings indicate that responses from the addressee causing the speaker to produce more gestures is not a plausible explanation for the effect of visibility on representational gestures.

The current study did not replicate previous findings that participants do not produce more gestures in the no screen condition than in the shoulder height screen condition. One possible explanation is that speakers in the current study were more considerate of the addressees due to both participants being students, and the ice breaker conversation at the beginning of the experiment. In the prior study (Barker & Kita, 2021), participants took on the role of speaker and the experimenter took on the role of the addressee. There was no ice breaker conversation. Speakers produced more gestures in the no screen and shoulder height screen condition than in the full screen condition, and there was no gesture height effect. It is plausible that speakers are more likely to take the perspective of their friends and peers than strangers. Members of the same community may be more likely to correctly interpret ambiguous utterances than those of distinct communities (Clark & Marshall, 1981). Speakers are also more likely to interpret ambiguous statements correctly in line with community expectations when speaking to a friend than a stranger (Gerrig & Littman, 1990). Furthermore, people are more accurate at identifying abstract stimuli when reading a description written by a friend rather than a stranger (Fussell & Krauss, 1989). In the current study, speakers may have felt more comfortable or more able to take the perspective of the addressee due to the ice breaker conversation and the status of the addressee as a peer.

The current study appears to contradict prior findings suggesting speakers are sensitive to whether the speaker's gestures can be seen (Mol et al., 2011). Over a webcam, speakers produced more gestures when the addressee could see the speaker. Speakers did not produce more gestures when the speaker could see the addressee. The finding appears to suggest that speakers are sensitive to whether their own gestures are useful for communication, and do not respond to the addressee's face as a cue. However, the findings of the previous study can be explained by the use of a web cam and the experimental instructions. The addressee's face may not have acted as a cue because eye contact was not possible. When eye contact could be made, speakers produced far more gestures when the speaker knew their gestures could be seen by the addressee (Mol et al., 2011). Furthermore, speakers may have altered their gesture production after the experimenter highlighted whether the addressee could see the speaker or not by displaying the addressee's point of view. It is plausible that showing the addressee's perspective reduced the cognitive cost of the speaker taking the perspective of the addressee. Therefore, speakers may have taken the perspective of the addressee and rely less on cues.

Previous free-dialogue designs may have not found an effect of visibility on representational gesture due to the stimuli used to elicit gesture. Studies that have not found an effect of visibility have used static spatial or abstract stimuli to elicit gesture (e.g. Bavelas et al., 2008), while studies that have found an effect of visibility have used dynamic spatial and motoric stimuli to elicit gesture (e.g. Alibali et al., 2001). If speakers produce more gestures to communicate information when discussing dynamic spatial and motoric topics, then the effect of visibility on gesture will be stronger. Gestures conveying spatial or motoric information result in gesture becoming more useful for communication (Hostetter, 2011), and speakers produce gestures more when describing spatial information than non-spatial information (Alibali, 2005; Rauscher et al., 1996). It is plausible that previous free-dialogue studies failed to find an effect of visibility on representational gesture because speakers were not motivated to produce gestures to benefit communication when describing static spatial or abstract topics. The current study may have found an effect of visibility in a free-dialogue conversational condition because the speakers were motivated to produce gestures to benefit communication when describing motoric and dynamic spatial topics.

Overall gesture rates were similar between the current study and previous free-dialogue studies (e.g. Bavelas et al., 2008). It is therefore possible that speakers produced communicative gestures at the same rate between the two studies. If speakers do not produce more gestures to communicate information when discussing dynamic spatial stimuli than static spatial stimuli, then the stimuli difference between the current study and previous us studies cannot explain the difference in visibility effects. However, it is possible that a higher proportion of the gestures in the current study were intended to communicate than in previous studies even if the overall gesture rate remain the same. It is possible that other differences between the studies could explain the differing proportions of communicative gestures while having the same overall gesture rate. For example, other differences between the two specific stimuli sets used in the studies could have affected the gesture rate. Differences between the populations the participants were drawn from could also have an effect. Furthermore, any other aspect of the interaction between the speaker and addressee could have affected the gesture rate. Further research should examine the use of communicative gestures directly across these different types of stimuli to establish that speakers were more motivated to produce gestures for communication when describing dynamic stimuli than spatial stimuli.

The current study did not replicate previous findings that speakers produce more gestures when conversing with a responsive addressee than a non-responsive addressee (e.g. Bavelas & Gerwing, 2011; Beattie & Aboudan, 1994; Jacobs & Garnham, 2007). That is, speaker's gesture production was unaffected by the addressee's responsive behaviour. One possible explanation is that addressees in the current experiment had little to contribute to the conversation. Speakers start with all the information and need to convey that information to the addressee. In one sided conversations, the addressee acts as a 'passive' addressee (Bavelas & Gerwing, 2011). Passive addressee's only need to convey their understanding of the speaker's message. It's possible that an 'active' addressee with more information to contribute may trigger speakers to produce more gestures.

Speakers may not have produced different rates of gesture in the responsive and non-responsive conditions because the addressee's behaviour matched the speakers' expectations in both conditions. The manipulation of addressee responsiveness differed from previous studies' manipulations. For example, Bavelas et al. (2008) did not have a non-responsive addressee in the equivalent non-responsiveness condition. The speaker talked to a tape recorder, instead of a non-responsive addressee. While it is possible that speakers produce more gestures the more responsive an addressee is, when no responses are provided, the speaker may change their behaviour to adapt to the complete lack of feedback. For example speaker's with distracted addressees perform worse on narrative tasks (Bavelas et al., 2000). In previous studies with silent addressees however, speakers produced a higher rate of gestures when talking to responsive addressees than non-responsive addressees (Beattie & Aboudan, 1994). The finding suggests that speakers still produce fewer gestures in response to silent addressees than responsive addressees. The findings of the current study may differ from previous studies because speakers expected responsive or non-responsive behaviours from the addressee and adapted to them. Speakers gesture more frequently when their expectations were consistent with addressees' behaviour (Kuhlen et al., 2012). For example, when they expected distracted addressees and addressees were indeed distracted. In the current study, speakers may have expected the addressee to be responsive or not after the practice session. This finding can explain why no effect of responsiveness was found; speakers adapted to the addressee's lack of responsiveness based off of their expectations.

Further research using the current study's data set could examine the content of the gestures to establish if the information expressed in gesture changes in response to cues from the addressee. It's plausible that the content of gesture changes when speakers are communicating information to the addressee, especially as more informative gestures benefit communication (Krason et al., 2021). It is possible that while speakers gestured more when the addressee's face was visible to the speaker, the gestures may have only been more informative to the addressee when the speaker's gestures were visible. Further research could establish if the speaker's gestures are more informative when the gestures are visible to the addressee, or if the addressee's face is visible to act as cue to the speaker.

4.5 Conclusion

To summarise, the current study is the first to provide evidence that speakers can both respond to cues from the addressee using heuristics and take the perspective of the addressee to communicate more effectively with the addressee. Speakers produced more representational gestures both in response to seeing the addressee's face, and the speaker's gestures being visible to the addressee. Furthermore, study is the first to provide evidence for the effect of visibility on representational gesture rate in a free-dialogue setting. The findings provided no evidence to suggest that the effect of visibility was due to the confounding of visibility and addressee responsiveness, as suggested by Bavelas and Healing (2013). However, the current study did not replicate previous findings that speakers produce more gestures when conversing with a responsive addressee than a non-responsive addressee.

Chapter 5

Foreground Gestures: The Effect of Foregrounding Gestures on Addressee Comprehension

Abstract

Speakers can design their gestures to communicate information with the addressee during conversation. One way that speakers can design their gestures to communicate information more effectively is to place gestures in the foreground of the interaction. Speakers can make gestures more visually prominent, fixate their gaze on their gestures, or refer to the gestures using speech to foreground the gestures. The current study investigated whether foreground and background gestures convey information to the addressee, and whether foreground gestures are more likely to convey information to the addressee than background gestures. Participants drew target events described by an actor in videos. Key target information was encoded only in gesture. Participants recalled a higher proportion of target information when gesture was encoded in visually prominent gestures and gestures indicated in speech than when the information was not encoded in gestures (Experiments 1 and 3). However, participants did not include significantly more target information when the actor produced gaze fixated gestures than when the actor did not produce a gesture (Experiment 2). The findings of the current study do not provide unequivocal evidence that placing gestures in the foreground benefits the addressee's comprehension. However, trends in the data suggest that making gestures visually prominent or referring to the gesture in speech may help the gesture to convey information to the addressee.

5.1 Introduction

Speakers can produce gestures alongside speech to convey information to an addressee. Co-speech gestures can take on forms which speakers can understand and take meaning from (Kendon, 2004). For example, speakers can produce pre-established conventionalised hand symbols (known as emblems (McNeill, 1992)) such as the 'ok' sign to convey a message to an addressee. Speakers can also produce gestures which depict an action, movement, object, shape, location, or metaphorically depict abstract concepts (such as weighing up to choices) to convey information (Iconic and Metaphorical gestures (McNeill, 1992)).

Co-speech gestures can successfully convey information to the addressee. Speakers communicate more effectively when producing gestures than when gestures are prohibited, suggesting that gestures significantly benefit the addressee's comprehension of the speaker's message (Dargue et al., 2019; Hostetter, 2011). More specifically, iconic gestures, metaphorical gestures and pointing gestures all benefit comprehension (Dargue et al., 2019). For example, addressees are better able to reproduce line drawings from descriptions given by speakers who gestured than by speakers who were prohibited from gesturing (Graham & Heywood, 1975). Children were better able to learn from teachers who produced gestures than teachers who did not produce gestures (Valenzeno et al., 2003). Addressees took in more information from adverts where a speaker was producing iconic gestures than audio only adverts, or tv adverts where information was conveying through pictures (Beattie & Shovelton, 2005). Overall, gestures provide a moderate benefit to the addressee's comprehension.

Co-speech gestures can convey information to varying degrees of effectiveness depending on the information being conveyed, the co-occurrence with speech, and the addressee (Hostetter, 2011). Gestures that convey information which was not present in speech were more effective at conveying information to the addressee than gestures which conveyed information which was redundant with speech (Hostetter, 2011)(c.f. Dargue et al., 2019). Gestures are more effective at conveying information when depicting motoric information such as actions or spatial dynamic information such as the movement of objects than when depicting abstract metaphorical information (Hostetter, 2011). Gestures are particularly suited to conveying spatial information rather than abstract information because gestures can efficiently convey action, shape, relative location of objects and the movement of objects (Hostetter, 2011). Addressees benefitted from seeing gestures conveying spatial and motoric gestures, but not from gestures conveying abstract information. Gestures are also particularly suited to conveying information to children (Hostetter, 2011). As children are less verbally proficient than adults, gesture may provide a larger role in compensating for difficult to comprehend messages in speech. For example, when the verbal message is complicated, children benefit significantly from the speaker's gestures (McNeil et al., 2000). Children benefit more than adults from speaker's gestures (Hostetter, 2011)(c.f. Dargue et al., 2019).

Speakers can also design their gestures to communicate more effectively with the addressee during speech. Speakers produce more gestures when an addressee could see the speaker's gestures (Alibali et al., 2001). Furthermore, speakers produce larger representational gestures when the addressee is visible (Bavelas et al., 2008). Speakers also placed gestures higher up in space when the addressee was visible (Holler et al., 2011). Speakers may produce gestures they wish to help communicate higher up in front of them and make them bigger. The larger, higher up gestures may be clearer and help the speaker convey information. Speakers will orient their gestures to move within a space they share with their addressee when speaking to a single addressee sitting diagonally to the side, or within a central space when speaking to two addressees sitting in front of the speaker (Özyürek, 2002). Speakers adapt their communication by presenting some spatial information in gesture rather than speech, suggesting speakers will present information in gesture when gesture is best suited to communicate that information to the addressee (Melinger & Levelt, 2004). Speakers take their addressee's viewpoint into account when producing pointing gestures (Winner et al., 2019). Speakers attenuate their gestures when retelling the story to an addressee they had already told the same story (Galati & Brennan, 2014; Jacobs & Garnham, 2007). Speakers also make communicative adjustments in response to addressee behaviours. Speakers' gestures become more precise, larger, or more visually prominent after feedback (Holler & Wilkin, 2011). Taken together, speakers adjust the production of gestures to increase the effectiveness of communication to the addressee.

5.1.1 Foreground Gestures

One way that speakers can design their gestures to communicate information more effectively is to place gestures in the foreground of the interaction. Speakers may design their gestures to attract the addressees attention and convey important information more clearly (Cooperrider, 2018; Streeck, 1993). Placing gestures in the foreground of the interaction can provide a clear indicator that the speaker produced the gesture to communicate information to the addressee (Streeck, 1993). Cooperrider (2018) defines foregrounded gestures as being in the forefront of both the speaker's and the addressee's awareness, as well as being in the foreground of the interaction. Gestures which are not placed in the foreground of the interaction are referred to as backgrounds gestures. In the current study, we investigated three key foregrounding behaviours. Speakers can place gestures in the foreground of the interaction by making the gesture larger and more visually prominent, by looking at their own gestures, and by referring to the gesture using speech.

5.1.1.1 *Visually prominent gestures*

Speakers can make gestures more visually prominent to put the gestures in the foreground of the interaction. Speakers may use more visually prominent gestures to signal the gestures use for communication (as suggersted by Cooperrider, 2018) and so that the gesture conveys information more clearly to the addressee (as suggested by Bavelas et al., 2008).

Speakers use more visually prominent gestures when the gestures are suited for conveying information to the addressee. For instance, speakers produce larger representational gestures when the addressee is visible (Bavelas et al., 2008). Speakers also placed gestures higher up in space when the addressee was visible (Holler et al., 2011). Speakers produced larger gestures when cooperating with the addressee during a task than when competing with the addressee. Speakers' gestures become larger and more visually prominent after feedback (Holler & Wilkin, 2011). The findings suggest that speakers use larger gestures whenever the speaker wishes to communicate information to the addressee via gesture.

There is no unequivocal empirical evidence to suggest that more visually prominent gestures convey information to the addressee more effectively than less visually prominent gestures. Previous studies have attributed speakers producing larger gestures to a desire to communicate more effectively with the addressee (e.g. Bangerter & Chevalley, 2007; Bavelas et al., 2008; Chu et al., 2014; Holler & Stevens, 2007; Holler & Wilkin, 2011; Hostetter et al., 2011). However, there are no findings to suggest that more visually prominent gestures communicate information more effectively either by drawing the attention of the addressee to the speaker's gestures, or by making the gesture clearer and more interpretable. The current study will examine if more visually prominent gestures benefit the addressee's comprehension of the speaker's message.

5.1.1.2 *Gaze fixated gestures*

Speakers may fixate their gaze at their own gestures to put the gestures in the foreground of the interaction. Speakers may look at their own gestures to signal the gestures use for communication and draw the addressee's attention to the gesture (Cooperrider, 2018; Streeck, 1993).

Speakers appear to look at their own gestures during conversations with an addressee (Enfield, 2009; C. Goodwin, 1986; Streeck, 1993). When speakers are fixating on their own gestures, speakers must be fully aware of their own gestures they are producing, which signals to the addressee that the gestures are intended to convey information. Despite numberous observations (e.g. Streeck, 1993), there is no unequivocal evidence that speakers gaze at their gestures more when trying to communicate information using gesture. For example, there is no established finding that spekers fixate more on their own gestures when the gestures are visible to the addressee.

Speakers fixating their gaze on their own gestures may affects the addressee's comprehension of the speaker's message. Addressee's attention is drawn to the speaker's gestures when speakers fixate on their own gestures (Gullberg & Kita, 2009). Furthermore, speakers fixating on their gestures did benefit addressee's comprehension of directional information encoded in the gestures (Gullberg & Kita, 2009). However, only an effect of gaze fixation on addressee's recall of direction information was tested. The finding has not been generalised to other types of information. Furthermore, the effect of gaze fixation on addressee comprehension was not reproduced when speech was controlled for, by using video editing to create artificial gaze fixations superimposed on the video with the same speech. In the initial experiment (Gullberg & Kita, 2009), gesture duration may have been confounded with gaze fixation as gestures were longer when the speaker fixated on the gestures than when the speaker did not fixate on the gestures. While the findings suggest the speakers fixating gaze on gestures benefits addressee comprehension, further work should expand the findings from only investigating directional information and address the potential confounding of gaze fixation, gesture duration, and speech production. The current study will examine if gaze fixated gestures benefit the addressee's comprehension of the speaker's message.

5.1.1.3 *Gestures indicated by demonstratives*

Speakers may use demonstratives to refer to their own gestures in speech to put their gestures in the foreground of the interaction. Speakers can refer to their gesture in speech using demonstratives (Guérin, 2015), that is, phrases such as 'like this' or 'like that'. Speakers can produce demonstratives to highlight and signal the gestures' use for communication.

Speakers use more demonstratives to refer to concurrent gestures when the gestures are suited for conveying information to the addressee. Speakers use more demonstratives when pointing to nearby objects than when pointing to objects further away (Bangerter, 2004). When the objects are closer, the gesture is easier to interpret so speakers will draw attention to the gesture. Furthermore, speakers use more demonstratives alongside gesture when the gestures are visible to the addressee than when gestures are not visible to the addresee (Bavelas et al., 2008; de Ruiter et al., 2012; Emmorey & Casey, 2001). Speakers are also especially likely to use demonstratives in association with gaze fixation behaviours when talking to children (Slonimska et al., 2015).

There is no unequivocal empirical evidence to suggest that gestures indicated by demonstratives convey information to the addressee more effectively than gestures which are not indicated by demonstrates. There are no findings to suggest that demonstratives referring to gestures draws the attention of the addressee to the speaker's gestures. The current study will examine if producing demonstratives alongside gestures benefits the addressee's comprehension of the speaker's message.

5.1.2 The Current Study

The current study investigated whether foregrounding gestures in the interaction helps the addressee comprehend the speaker's message. In the current study, we manipulated how gestures were foregrounded in three different ways, over three different experiments. In Experiment 1, we investigated the effect of making gestures visually prominent on addressee comprehension. In Experiment 2, we investigated the effect of the speaker fixating their gaze on their gestures on addressee comprehension. In Experiment 3, we investigated the effect of the speaker producing demonstratives alongside their gestures on addressee comprehension.

We investigated the effect of foregrounding gestures on addressee comprehension by examining two questions. First, do foreground and background gestures convey information at all? Secondly, are foreground gestures more likely to convey information to the addressee than background gestures? We hypothesised that foreground gestures (e.g. visually prominent) are more likely to convey information to the addressee than background gestures (i.e. non-visually prominent). Furthermore, we hypothesised that both foreground and background gestures should convey information to the addressee to some degree (but foreground gestures more than background gestures).

To test our hypothesise, we tested the addressee's recall of target information expressed in gestures in the three experiments. Participants saw videos of an actor describing cartoons. For each cartoon, several target events could potentially contain target information. The target information was encoded solely in gesture. Participants would draw a picture of each target event from the cartoon based on the actor's description. For each target event drawing, we checked if the target information (from the gesture) was included in the drawing. For each experiment, we examined the addressee's recall of target information in three conditions. In the first condition, target information was encoded by foreground gestures. In the second condition, target information was encoded by background gestures. In the third condition, target information was not encoded because no gestures were produced (the gesture was *absent*). To test the first hypothesis that foreground and background gestures convey information to the addressee, we compared participant's recall in the foreground and background gesture conditions to the absent gesture condition. To test the second hypothesis that foreground gestures are more likely to convey information to the addressee than background gestures, we compared participant's recall between the foreground gesture and background gesture conditions. These comparisons were made for all three foregrounding behaviours being examined: Visual prominence, gaze fixations, and gestures indicated by demonstratives.

We made the same predictions for all three foregrounding behaviours across the three experiments. We predict that the addressees will recall more target information in the foreground condition than the background condition. Furthermore, we predict that the addressees will recall more target information in the foreground and background conditions than in the absent (no gesture) condition.

5.2 Experiment 1

5.2.1 Introduction

We investigated the effect of visually prominent gestures on addressee comprehension to test two predictions. First, if both visually prominent and non-visually prominent gestures convey information to the addressee compared to when no gesture is produced. Second, if visually prominent gestures convey information more effectively than non-visually prominent gestures.

5.2.2 Methods

5.2.2.1 *Participants*

Participants were recruited in two ways. One set of participants were first year Psychology undergraduate students from the University of Warwick who took part for course credit. The other set of participants were recruited using the department's participant recruitment system in exchange for a chance to win a voucher. Participants were only included in analysis if they spoke English as their first language, or if they self-reported their fluency with English as excellent. Furthermore, participants were only included in analysis if they successfully submitted their drawings. 27 participants took part in the experiment (Gender, 26 women, 1 man. Age, M = 18.81, SD = 1.08). The number of participants recruited was based on a previous gesture comprehension in which speakers had to produce drawings (Theron-Grimaldi, 2021). We used the software program G*Power to conduct a power analysis. Our goal was to obtain .95 power to detect an effect size smaller than those found by a previous similar experiment (Theron-Grimaldi, 2021), (effect size f = .81 / partial Eta Squared = .397) at the standard .05 alpha error probability. We estimated a medium effect size (f) of .26 for our effect.

5.2.2.2 *Design*

The independent variable was gesture type (V*isually prominent, non-visually prominent, absent*). The dependant variable was the inclusion or absence of target information in each drawing (*Present, Absent*).

5.2.2.3 *Materials*

During the experiment, addressee's watched premade stimuli videos. In each video, the actor described a scene adapted from a Tweety and Sylvester film entitled "Canary Row". Participants saw eight videos featuring 30 target events. The videos target events were always presented in a fixed order. Each target event included some target information which would only be encoded in gesture (or not at all in the gesture absent condition). Each target event was either A) accompanied by a visually prominent gesture (visually prominent condition), B) accompanied by a non-visually prominent gesture (non-visually prominent condition), or C) was not accompanied by a gesture (gesture absent condition). Across the eight videos, participants saw ten target events accompanied by a visually prominent gesture, ten target events accompanied by a non-visually prominent gesture, and ten target events not accompanied by a gesture. The conditions rotated for each target event so that for every three target events, one target event was assigned to each condition (i.e. the first three target events a participant saw might be a visually prominent gesture event, then a non-visually prominent gesture event, and then an absent gesture event). Therefore, each of the eight videos contained at least one target event for all three conditions. Three sets of the eight videos were created for counterbalancing purposes. We counterbalanced condition order using the Latin square technique. The first participant saw a visually prominent gesture event first, the second participant saw a non-visually prominent gesture event first, and the third participant saw an absent gesture event first. Thus, each target event appeared equally often in each condition. The actor in the videos wore a mask obscuring their mouth. The audio was recorded separately and edited into the videos so that speech was identical across conditions.

Each target event had one piece of target information which would only be encoded in gesture. The description of each 'event' was accompanied by either a visually prominent gesture, a non-visually prominent gesture, or no gesture at all. In the *visually prominent* condition, the actor produced a large representational gesture and used the space from the actor's waist to their head. In the non-*visually prominent* condition, the actor produced a small gesture in the periphery of gesture space if possible, such as the actor's lap. For events in the *absent* condition, the actor did not produce a gesture (the target information was not encoded). Image examples of the three conditions can be found in Appendix 4, figures 1, 2 and 3 respectively. The full set of stimuli can be accessed here: https://osf.io/u2cye/?view_only=c4b1f3f5747e48308b2e0521d2101c85

To elicit the drawings, participants were shown two pictures for each 'event' they were asked to draw. The pictures were taken directly from 'Canary Row'. One picture would act as a 'before' picture, while the other would act as an 'after' picture. Participants were asked to draw on a premade set of numbered squares they could print off.

5.2.2.4 *Procedure*

Participants first were briefed about the task on Qualtrics. Participants prepared to take part in the experiment by printing off the drawing sheet with numbered squares. Once participants were ready to start the experiment, they were sent to a University of Warwick web application to take part in the drawing trials. Participants would view two videos of the actor describing scenes from the cartoon. After both videos had finished, participants would then see a pair of images from the cartoon and be asked to draw the event that took place between the events depicted in the two images. Participants were not allowed to proceed to the next drawing event until at least thirty seconds had passed. When the participants had finished the drawings for the first two videos, the participants then watched the next two videos. This procedure continued until the participants had watched all eight videos and completed all thirty event drawings.

After completing all the videos, participants were asked to scan in the drawings and upload them using the web application. Participants were then returned to the Qualtrics form to fill out demographic and English language fluency questions.

5.2.2.5 *Drawing coding*

Each of the participant's thirty drawings were coded by a research assistant blind to the conditions. For each drawing, the coder marked the target information as being present or absent in the drawing according to the coding scheme (Appendix 4 – including examples of present and absent target information drawings).

5.2.3 Results

5.2.3.1 *Data analysis*

The binary dependant variable (Target information present or absent) was analysed using a mixed effect logistic regression analysis. We used a maximal random effects structure (Barr et al., 2013) for the model by including all random slopes, random intercepts, and the covariance between the two for participants and target events. If the model obtains a singular fit, a second model will be reported. The second model will be the most maximal model without producing a singular fit. We analysed the data using R software for statistical analyses (R Development Core Team, 2011), with the lme4 package (Bates et al., 2015). Any main effects of gesture will be tested by examining the beta estimates of the effects (β) in a comparison of the three gesture type conditions.

We used likelihood ratio tests to compare models with and without the gesture type variable to establish a main effect of gesture type. For each model, we calculated the marginal and conditional R^2 using the *piecewise*SEM package (Nakagawa & Schielzeth, 2013). The marginal R^2 represents the variance explained by the fixed factors, while the conditional R^2 represents the variance explained by both fixed and random factors. If a model is obtained with a singular fit, the R^2 cannot be calculated.

5.2.3.2 *Maximal model*

Participant's recall of the target information was entered into a maximal logistic mixed effects model with gesture type as a within-subject variable (See figure 1 for descriptive statistics). Participants and target events were included as random factors. The main effect of gesture type was significant, $\chi^2(2) = 8.11$, $p = .017$. The model obtained a singular fit so we cannot estimate the variance explained by the model.

To test the nature of the main effect of gesture type, we examined the beta estimates in a comparison of the three gesture type conditions. To compare the visually prominent and non-visually prominent conditions to the absent condition, the gesture absent condition was used as the reference level. To compare the visually prominent condition to the non-visually prominent condition, the non-visually prominent condition was used as the reference level. The Participants recalled more target information in the visually prominent gesture condition than the absent gesture condition ($\beta = 0.93$, $SE = 0.33$, $p = .004$). The other pairwise comparisons were not significant: visually prominent vs non-visually prominent ($\beta = 0.31$, $SE = 0.29$, $p = .284$), non-visually prominent vs absent ($\beta = 0.62$, $SE = 0.34$, $p = .066$).

5.2.3.3 *Simplified model*

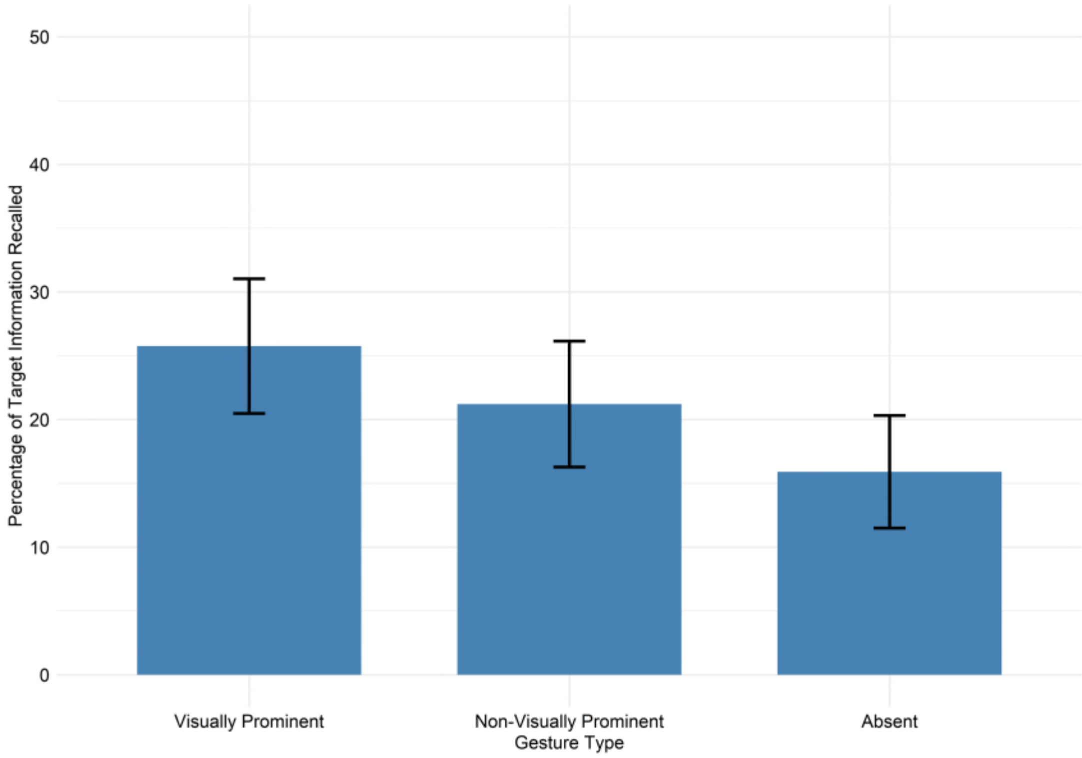
As the maximal model obtained a singular fit, we will also report a simplified model. The simplified model included random intercepts, but not random slopes. We compared the maximal and simplified models. The maximal model with random slopes did not fit the data significantly better than the simplified model without random slopes $\chi^2(10) = -2.08$, $p = .996$.

In the simplified model, the main effect of gesture type was significant, $\chi^2(2) = 8.68$, $p = .013$. The model explained approximately 21% of the variance in participants recall of target information (marginal $R^2 = .01$, conditional $R^2 =.21$).

Pairwise comparisons yielded the same results as in the maximal model: Visually prominent vs absent ($\beta = 0.70$, $SE = 0.24$, $p = .003$), visually prominent vs non-visually prominent ($\beta = 0.23$, $SE = 0.22$, $p = .306$), non-visually prominent vs absent ($\beta = 0.47$, $SE = 0.24$, $p = .054$).

Figure 1

The percentage of target information drawn in the visually prominent, non-visually prominent, and absent gesture type conditions in Experiment 1. The error bars represent 95% confidence intervals.



5.2.4 Discussion

There are three key findings. First, participants included the target information more often in the visually prominent gesture condition than the absent gesture condition. Second, the proportion of target information recalled in the visually prominent gesture condition did not significantly differ from the proportion of target information recalled in the non-visually prominent condition. Third, the proportion of target information recalled in the non-visually prominent gesture condition did not significantly differ from the proportion of target information recalled in the absent gesture condition. The findings indicated that visually prominent gestures convey information to the addressee, but the findings do not provide evidence that non-visually prominent gestures convey information to the addressee. Furthermore, the findings do not provide evidence that visually prominent gestures were more effective at communicating information to the addressee than non-visually prominent gestures. The findings do not support the claim that foreground gestures are more likely to convey information to the addressee than background gestures.

5.3 Experiment 2

5.3.1 Introduction

We investigated the effect of gaze fixated gestures on addressee comprehension to determine two outcomes. First, if both gaze fixated and non-gaze fixated gestures convey information to the addressee compared to when no gesture is produced. Second, if gaze fixated gestures convey information more effectively than non-gaze fixated gestures.

5.3.2 Methods

5.3.2.1 *Participants*

Participants were recruited in the same way as in Experiment 1. 27 participants took part in the experiment (Gender, 24 women, 2 men, 1 Non-binary. Age, M = 19.11, SD = 0.93).

5.3.2.2 *Design*

The design was the same as the design of Experiment 1 except that the within-subject independent variable was gesture type (*Gaze fixated gesture, non-Gaze fixated gesture, Absent*).

5.3.2.3 *Materials*

The materials were created in the same way as for Experiment 1, but gaze fixations were manipulated instead of visual prominence. Each target event was either A) accompanied by a gaze fixated gesture (gaze fixated condition), B) accompanied by a non-gaze fixated gesture (non-gaze fixated condition), or C) was not accompanied by a gesture (gesture absent condition).

Each target event had one piece of target information which would only be encoded in gesture. The description of each 'event' was accompanied by either a gaze fixated gesture, a non-gaze fixated gesture, or no gesture at all. In the *gaze fixated* condition, the actor produced and looked at a representational gesture. The speaker began fixating on the gesture at the beginning of the stroke phase and ends fixating on the gesture at the end of the stroke phrase. Gestures were not timed so the duration of the fixation would differ between target events. In the non-*gaze fixated* condition, the actor continued to look toward the camera and produced a representational gesture. For events in the *absent* condition, the actor did not produce a gesture (the target information was not encoded). Image examples of the three conditions can be found in Appendix 4, figures 4, 5 and 3 respectively. The full set of stimuli can be accessed here: https://osf.io/u2cye/?view_only=c4b1f3f5747e48308b2e0521d2101c85

5.3.2.4 *Procedure*

The procedure was the same as Experiment 1.

5.3.2.5 *Drawing coding*

drawings were coded in the same way as in Experiment 1.

5.3.3 Results

5.3.3.1 *Data analysis*

The data was analysed in the same way as the data for Experiment 1

5.3.3.2 *Maximal model*

Participant's recall of the target information was entered into a maximal logistic mixed effects model with gesture type as a within-subject variable (See figure 2 for descriptive statistics). Participants and target events were included as random factors. The gesture absent condition was used as the reference level for the model. The main effect of gesture type was *not* significant, $\chi^2(2) = 3.12$, $p = .210$. The model obtained a singular fit so we cannot estimate the variance explained by the model.

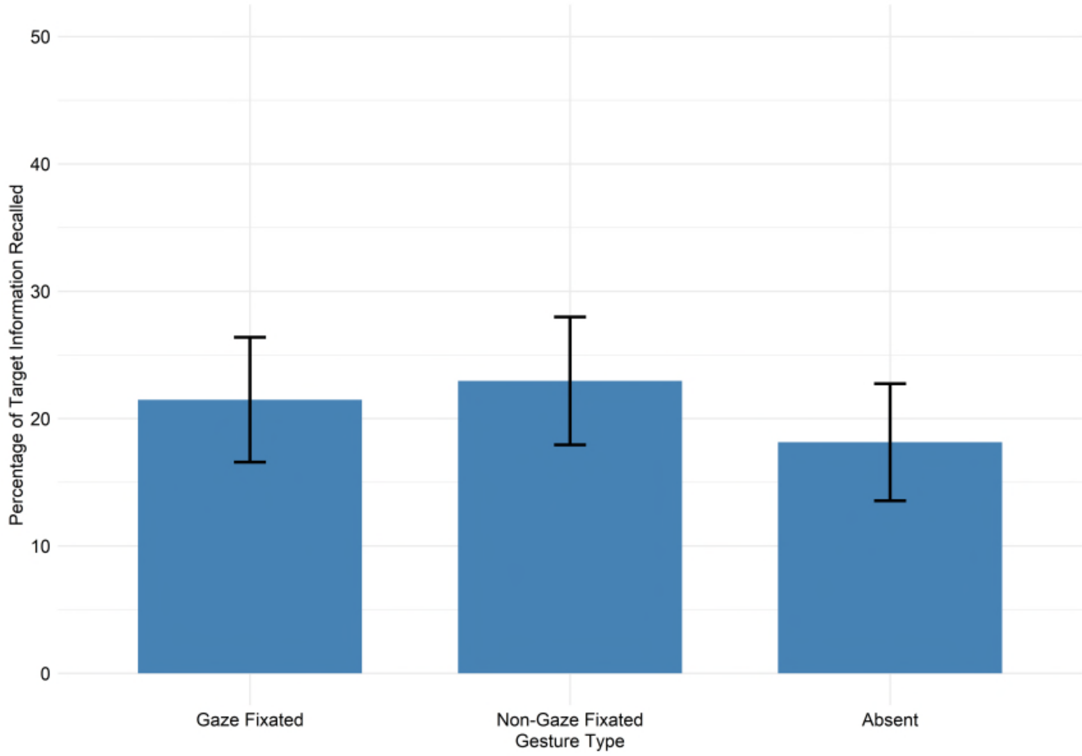
5.3.3.3 *Simplified model*

As the maximal model obtained a singular fit, we will also report a simplified model. The simplified model included random intercepts, but not random slopes. We compared the maximal and simplified models. The maximal model with random slopes did not fit the data significantly better than the simplified model without random slopes $\chi^2(10) = -7.27$, $p = .700$.

In the simplified model, the main effect of gesture type was *not* significant, $\chi^2(2) = 1.18$, $p = .554$. The model explained approximately 13% of the variance in participants recall of target information (marginal $R^2 = .001$, conditional $R^2 = .13$).

Figure 2

The percentage of target information drawn in the gaze fixated, non-gaze fixated, and absent gesture type conditions in Experiment 2. The error bars represent 95% confidence intervals.



5.3.4 Discussion

In Experiment 2, the proportion of target information in either the gaze fixated or non-gaze fixated gesture condition did not significantly differ from the proportion in the absent gesture condition. Furthermore, the proportion of target information in the gaze fixated gesture condition was not significantly different than in the non-gaze fixated condition. The findings do not provide evidence that gaze fixated gestures were more effective at communicating information to the addressee than non-visually prominent gestures. Furthermore, the findings do not provide evidence that either gaze fixated or non-gaze fixated convey information to the addressee. The findings do not support the claim that foreground gestures are more likely to convey information to the addressee than background gestures. The findings also do not support the claim that foreground or background gestures convey information to the participant.

5.4 Experiment 3

5.4.1 Introduction

We investigated the effect of gestures indicated by demonstratives on addressee comprehension to determine two outcomes. First, if both gestures indicated by demonstratives and gestures not indicated by demonstratives convey information to the addressee compared to when no gesture is produced. Second, if gestures indicated by demonstratives convey information more effectively than non-gestures indicated by demonstratives.

5.4.2 Methods

5.4.2.1 *Participants*

Participants were recruited in the same way as in Experiment 1. 31 participants took part in the experiment (Gender, 24 women, 4 men, 1 non-binary, 2 preferred not to say. Age, M = 18.87, SD = 0.85).

5.4.2.2 *Design*

The design was similar to the design of Experiment 1. The within-subject independent variable was gesture type (Gesture with demonstrative, *Gesture without demonstrative, Gesture absent*).

5.4.2.3 *Materials*

The materials were created in a similar way as for Experiment 1 but gesture accompanying demonstratives were manipulated instead of visual prominence. Each target event was either A) accompanied by a gesture indicated by a demonstrative (gesture with demonstrative condition), B) accompanied by a gesture which is not indicated by a demonstrative (gesture without demonstrative condition), or C) was not accompanied by a gesture (gesture absent condition).

Furthermore, six sets of the eight videos were created for counterbalancing purposes (rather than three sets). The new three sets were created by editing the audio so that gestures could be indicated by a demonstrative or not indicated by a demonstrative. We created three original recordings. In the original recordings, demonstratives were produced to accompany gestures (but not in target events where no gesture was produced for the gesture absent condition). We created three videos following the same latin square structure as in the previous experiments (i.e. three sets of videos with the following order: gesture with demonstrative condition, gesture without demonstrative condition, gesture absent) The starting condition was different for each of the three sets. We created a further three videos following a new Latin square order (I.e. three sets of videos with the following order: gesture without demonstrative condition, gesture with demonstrative condition, absent). The starting condition was different for each of the three sets.

The audio was edited to remove demonstratives from the target events intended to be non-demonstrated. To avoid a long pause in speech during the gesture, the experimenter edited the audio so the speech before was slightly later, and the speech after was slightly earlier. The speech edits only affected the speech just before and after each target event and did not affect the speech co-occurring with any other gestures. Furthermore, the actor wore a mask so there were no lip-syncing issues.

Each target event had one piece of target information which would only be encoded in gesture. The description of each 'event' was accompanied by either gesture with a demonstrative, a gesture without a demonstrative, or no gesture at all. In the *gesture with demonstrative* condition, the actor produced a representational gesture and uttered a demonstrative at the same time. In the *gesture without demonstrative* condition, the actor produced a representational gesture. For events in the a*bsent* condition, the actor did not produce a gesture (the target information was not encoded). The full set of stimuli can be accessed here: https://osf.io/u2cye/?view_only=c4b1f3f5747e48308b2e0521d2101c85

5.4.2.4 *Procedure*

The procedure was similar to the procedure of Experiment 1, except participants were allocated to one of six counterbalance orders (rather than 3).

5.4.2.5 *Drawing coding*

In Experiment 3, drawings were coded in the same way as in Experiment 1.

5.4.3 Results

5.4.3.1 *Data analysis*

The data for Experiment 3 was analysed in the same way as the data for Experiment 1

5.4.3.2 *Maximal model*

Participant’s recall of the target information was entered into a maximal logistic mixed effects model with gesture type as a within-subject variable (See figure 3 for descriptive statistics). Participants and drawings were included as random factors. The main effect of gesture type was significant, $\chi^2(2) = 6.31$, $p = .043$. The model obtained a singular fit so we cannot estimate the variance explained by the model.

To test the main effect of gesture type, we examined the beta estimates in a comparison of the three gesture type conditions. To compare the gesture with demonstrative and gesture without demonstrative conditions to the absent condition, the gesture absent condition was used as the reference level. To compare the gesture with demonstrative condition to the gesture without demonstrative, the gesture without demonstrative condition was used as the reference level. Participants recalled more target information in the gesture with demonstrative condition than the absent gesture condition ($\beta = 0.87$. $SE = 0.32$, $p = .007$). The other pairwise comparisons were not significant: gesture with demonstrative vs gesture without demonstrative ($\beta = 0.33$, $SE = 0.27$, $p = .226$). Gesture without demonstrative vs absent gesture condition ($\beta = 0.54$. $SE = 0.33$, $p = .099$).

5.4.3.3 ***Simplified model***

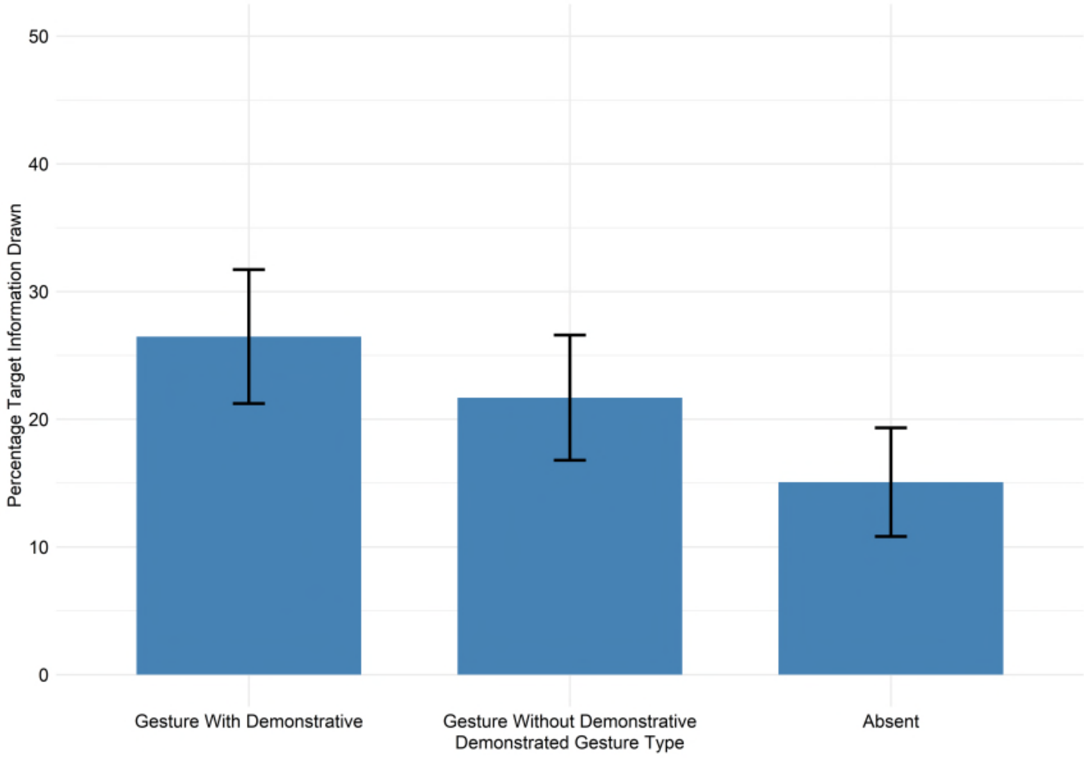
As the maximal model obtained a singular fit, we will also report a simplified model. In the simplified model, we used a random effects structure for the model including random intercept variation, but not random slope variation. We compared the maximal and simplified models. The maximal model with random slopes did not fit the data better than the simplified model without random slopes $\chi^2(10) = -1.89$, $p = .997$.

In the simplified model, the main effect of gesture type was significant, $\chi^2(2) = 12.91$, $p = .002$. The model explained approximately 19% of the variance in participants recall of target information (marginal $R^2 = .02$, conditional $R^2 =.19$).

To test the main effect of gesture type, we examined the beta estimates in a comparison of the three gesture type conditions. The same reference levels were used for the simplified model comparisons as for the maximal model comparisons. The results differed from the maximal model. Participants recalled more target information in the gesture with demonstrative condition than the absent gesture condition ($\beta = 0.84$. $SE = 0.23$, $p < .001$). Participants did *not* recall significantly different proportions of information in the gesture with demonstrative condition than in the gesture without demonstrative condition ($\beta = 0.32$. $SE = 0.22$, $p = .142$). Contrary to the maximal model, Participants did recall significantly more information in the gesture without demonstrative condition than in the absent condition (β = 0.51. *SE* = 0.24, *p* = .031).

Figure 3

The percentage of target information drawn in the gesture with demonstrative , gesture without demonstrative, and absent gesture type conditions in Experiment 3. The error bars represent 95% confidence intervals.



5.4.4 Discussion

In Experiment 3, the maximal model and the simplified model found different effects. We will discuss the simplified model as the maximal model had a singular fit and was not a significant improvement over the simplified model when the models were compared. There are three key findings. First, participants included the target information more often in the demonstrated gesture condition than the absent gesture condition. Second, the proportion of target information recalled in the demonstrated gesture condition did not significantly differ from the proportion of target information recalled in the non-demonstrated condition. Third, participants included the target information more often in the non-demonstrated gesture condition than in the absent gesture condition in the simplified model, but the effect was not significant in the maximal model. The findings indicated that both demonstrated and non-gestures indicated by demonstratives convey information to the addressee. The findings do not provide evidence that gestures indicated by demonstratives were more effective at communicating information to the addressee than non-gestures indicated by demonstratives. The findings overall do not support the claim that foreground gestures are more likely to convey information to the addressee than background gestures. However, the findings of the maximal model suggest that only foreground convey information to the addressee.

5.5 General Discussion

The current study investigated whether placing gestures in the foreground of the interaction benefits the addressee's comprehension of the speaker's message. Overall, all three Experiments failed to support the claim that placing gestures in the foreground of the interaction rather than the background benefits the addressee's comprehension. In all three Experiments, the proportion of target information recalled in the foreground gesture condition did not significantly differ from the proportion of target information recalled in the background gesture condition. The findings of visually prominent gestures and gestures indicated by demonstratives were similar. In both Experiments 1 and 3, participants included the target information more often when the actor produced foreground gestures than when the actor did not produce any gestures. Furthermore, in both Experiments 1 and 3, participants did not include the target information significantly more often when the actor produced background gestures than when the actor did not produce any gestures (except in the simplified model of Experiment 3). The findings of gaze fixated gestures differed from the other two foregrounding behaviours. In Experiment 2, participants did not include significantly more target information when the actor produced gaze fixated or non-gaze fixated gestures than when the actor did not produce a gesture. Taken together, only the findings in Experiment 1 and 3 partially supports the claim that placing gestures in the foreground of the interaction benefits the addressee's comprehension of the speaker's message.

The trends in the data for Experiments 1 and 3 (visual prominence and demonstratives) but not Experiment 2 (gaze fixation) support the claim that placing gestures in the foreground of the interaction benefits the addressee's comprehension of the speaker's message. For both visually prominent gestures and gestures indicated by demonstratives, the trend in the data suggests that foreground gestures lead to higher rates of target information being included in participants drawings than gestures in the background. Furthermore, the trend suggests that both foreground and background gestures can convey information to the addressee (more than in the absent gesture condition). The trends observed in the current study suggest that gestures which are more visually prominent or accompanied by a demonstrative are more effective at communicating information than gestures in the background of the interaction.

The current study is the first to investigate the effect of gesture visual prominence on addressee's comprehension. Previous studies have assumed that larger gestures are an indicator that the speaker intends to communicate informaiton to the addressee (e.g. Bavelas et al., 2008; Holler & Wilkin, 2011). The findings did not provide evidence for the claim that more visually prominent gestures communicate more effectively than non-visually prominent gestures, although only visually prominent gestures significantly conveyed information to the addressee. The findings of the current study, and the trend visible in the data suggest that visually prominent gestures speakers may use visually prominent gestures to convey information to the adressee. Further work is needed to establish the role visually prominent gestures play in communicating information and signaling the speaker's intention to communicate.

The current study is also the first to investigate the effect of producing demonstratives alongside gestures on addressee's comprehension. Previous studies have assumed that gestures which are accompanied by demonstratives are an indicator that the speaker intends to communicate informaiton to the addressee using gesture (e.g. Bavelas et al., 2008; de Ruiter et al., 2012; Emmorey & Casey, 2001). The findings did not provide evidence for the claim that gestures which are accompanied by demonstratives communicate information more effectively than gestures which are not accompanied by demonstratives, although only gestures accompanied by demonstratives (in the maximal model) significantly conveyed information to the addressee. As the trend in the data suggests that gestures accompanied by demonstratives do benefit communication more than gestures which are not accompanied by demonstratives, further research is needed to establish how demonstratives benefit communication.

The current study adds to the previous research on the effect of speakers fixating their gaze on their gestures. Speakers can fixate their gaze on their own gestures to draw the addressee's attention to gesture (Gullberg & Kita, 2009). However, it was unclear if the addressees benefitted from speakers fixating on the gestures, as gesture duration was confounded with the effect of gaze fixation (Gullberg & Kita, 2009). Furthermore, it was unclear if the effect of gaze fixation on addressee comprehension generalised beyond directional information. The current study cannot support the claim that the effect of gaze fixation on addressee comprehension generalised beyond directional information.

The current study elicited lower levels of target information recall than previous studies. In a previous study where participants were asked to draw events containing target information, target information was included in over 50% of the drawings (Theron-Grimaldi, 2021). In this previous study, gesture size was not controlled for, and no demonstratives or gaze fixations were used by the speaker. In the current study, target information was never included in more than 28% of the drawings. The key difference in methodologies which may have caused this disparity in findings is the current study was conducted online, while the prior study was conducted in person. It is plausible that participants were more motivated to attend to the stimuli in person with an actor present. If participants in the current study were less motivated to attend to the stimuli, then the target information would be more difficult to recall. Furthermore, information expressed in gesture may have been missed leading to a reduced effect of gesture type on addressee comprehension.

5.6 Conclusion

The findings of the current study do not provide unequivocal evidence that placing gestures in the foreground of the communicative interaction benefits the addressee's comprehension. The trends in the data suggest that making gestures visually prominent or referring to the gesture in speech (but not looking at the gesture) may help the gesture to convey information to the addressee. Further research is necessary to establish if foregrounding behaviours benefit communication.

Chapter 6

General Discussion

In the current thesis, we have examined how speakers use gestures to communicate information to the addressee. Speakers can use gesture to convey information to an addressee (Kendon, 2004). Speakers communicate information to an addressee more effectively when producing gestures than when gestures are prohibited (Hostetter, 2011). Furthermore, speakers produce gestures differently depending on the context to communicate more effectively to the addressee. For example, speakers produce more gestures when the gestures are visible to the addressee because the gestures can convey information to the addressee (Alibali et al., 2001). By placing gestures in the "foreground" of an interaction, gestures are thought to be more effective at communicating information to the addressee (Cooperrider, 2018; Streeck, 1993).

In the current Chapter, we have considered the findings together to answer the three key research questions. First, we have investigated what mechanism can underly audience design behaviours including gestural audience design behaviours. Second, we have investigated which factors can affect co-speech gesture including gesture visibility, face visibility, addressee responsiveness and stimuli topic. Third, we have investigated if gestural audience design behaviours result in addressees' better comprehending the speaker's message. More precisely, we have investigated if placing gestures in the foreground of an interaction leads to increased addressee comprehension. We have also considered the implication the findings have for the literature, both for audience design research and for non-verbal communication research. We have proposed further research needed to expand on the findings in the current thesis.

6.1 The research questions

6.1.1 The first research question

The first research question is what mechanisms are responsible for producing audience design behaviours. The perspective taking account of audience design assumes that speakers take the perspective of the addressee to infer if gestures are visible (and therefore useful for communication). A cue-based heuristic mechanism had been put forward as a plausible mechanism but had not been empirically supported in the prior literature. The perspective taking explanation had not previously been compared to the cue-based heuristic explanation in an experimental study. We compared the cue-based heuristic and perspective taking mechanisms directly in a series of experiments.

To answer the first research question, we have conducted a series of experiments investigating the effect of visibility on gesture. In both chapters 3 and 4, speakers and described stimuli and spontaneously produced gestures to an addressee. In addition to the conventional manipulation of visibility, we introduced a novel condition in which a shoulder height screen was placed between the speaker and the addressee to prevent gestures being seen by the addressee while allowing the speaker and addressee to see each other's faces. By separating when the speaker's gestures and the addressee's face were visible, we were able to create distinct predictions for the perspective taking and cue-based heuristic mechanisms.

In Chapters 3 and 4, we found that speakers can design communicative behaviours both by taking the perspective of the addressee, and by heuristically responding to cues from the addressee. In Chapter 3, speakers produced more representational gestures in response to seeing the addressee's face but not in response to the addressee being able to see the speaker's gestures. The findings are consistent with the cue-based heuristic view of the effect of visibility on gesture. The findings were not consistent with the perspective taking view. In the third experiment of Chapter 3, speakers did not produce more representational gestures in the chin upward screen condition (where gestures were visible but not the addressee's face) than in the full screen condition. The findings suggest that the addressee's face acts as a cue for speakers to produce more gestures for the addressee. In Chapter 4 however, speakers produced more representational and interactive gestures when the speaker's gestures and the addressee's face were visible than when only the addressee's face was visible. Furthermore, speakers produced their representational gestures higher up when the gestures could be seen by the addressee. The findings in Chapter 4 suggests that speakers were able to take the perspective of the addressee and infer whether gestures would be useful for communication. Speakers still produced more representational gestures when the addressee's face was visible than when the addressee's face was not visible, suggesting that the speakers face was still acting as a cue for the speaker to produce more gestures. Taken together, the findings of Chapter 3 and 4 suggest that speakers can produce audience design behaviours both in response to cues from the addressee, and by taking the perspective of the addressee and inferring how to adjust communicative behaviours.

The disparity in findings between Chapters 3 and 4 can be plausibly explained by different levels of speaker motivation. In Chapter 3, speakers described stimuli to the experimenter. It is possible that speakers assumed the knowledge state of the experimenter. Speakers may have believed the experimenter was already knowledgeable about the cartoons, and that the experimenter was not under pressure to correctly recall the story afterwards. Therefore, speakers may have been less motivated to communicate information effectively to the experimenter. In Chapter 4, speakers described stimuli to naïve participants. It is plausible that speakers believed it was important to communicate the information to the fellow participant. Speakers may have believed the student addressee was ignorant about the cartoons, and that the addressee was under pressure to correctly recall the story afterwards. If speakers believed it was important to clearly convey the information to the addressee, then speakers may have been more motivated to take the perspective of the addressee to ensure effective communication.

If speakers are more motivated to communicate effectively, then speakers may put more resources into taking the perspective of the addressee to help communicate information more effectively. It is plausible that speakers may rely on cue-based heuristics for audience design behaviours until speakers are sufficiently motivated to take the perspective of the addressee. Participants are more likely to accurately take the perspective of an addressee when there is a strong motivation to accurately guess how the addressee will interpret a phrase (Epley et al., 2004). Thus, speakers who are motivated to take the perspective of the addressee (such as when the addressee is a fellow student who may be asked about the cartoons later) will attribute more cognitive resources to perspective taking. Speakers may have needed to have been sufficiently motivated to take the perspective of the addressee as perspective taking can be cognitively demanding (Rossnagel, 2000). The suggestion that speakers only take the perspective of the addressee when motivated to do so fits with predictions of the interactive-alignment model (Pickering & Garrod, 2004). According to Pickering and Garrod, (2004, p. 179), "performing inferences about common ground is an optional strategy that interlocutors employ only when resources allow. Critically, such strategies need not always be used, and most "simple" (e.g., dyadic, non-didactic, non-deceptive) conversation works without them most of the time".

It is also plausible that speakers will be more likely to take the addressee when the addressee is known to themselves. Members of the same community may be more likely to correctly interpret ambiguous utterances than those of distinct communities (Clark & Marshall, 1981). Speakers are more likely to interpret ambiguous statements correctly in line with community expectations when speaking to a friend than a stranger (Gerrig & Littman, 1990). People are more accurate at identifying abstract stimuli when reading a description written by a friend rather than a stranger (Fussell & Krauss, 1989). In Chapter 4, speakers may have felt more comfortable or more able to take the perspective of the addressee due to the ice breaker conversation and the identity of the addressee as a peer. However, the suggestion that speakers are more likely to take the perspective of peers and friends may conflict with the idea that if communicative partners share a more similar representation, then there is a reduced need to access a model of the addressee (Pickering & Garrod, 2004). According to the interactive-alignment model (Pickering & Garrod, 2004), speakers and addressees align linguistic representations during interactions. The more aligned the representations become, the more effective communication becomes. Furthermore, the more aligned representations become, perspective taking becomes less and less necessary. Therefore, it is plausible that as members of the same cohort, studying the same subject and possibly having the same friends, the speakers and addressees may already have some form of shared representation. If this suggestion is correct, then being peers with the addressee may have reduced the need for perspective taking.

Taken together, speakers can design communicative behaviours both heuristically responding to cues from the addressee as a default, and also by taking the perspective of the addressee when the speaker is motivated to do so.

6.1.2 The second research question

The second research question is what competing factors affect gesture production during conversation. Several factors have potentially been confounded in prior gesture studies including mutual visibility of gesture, mutual visibility of faces, the addressee's responsive behaviour (both visual and verbal), and the topic of discussion. Several studies have found an effect of visibility on representational gesture rate (e.g. Alibali et al., 2001), while several other studies have not (e.g. Bavelas et al., 2008; see Bavelas & Healing, 2013, for a review). The difference has been attributed to the different behaviour of the addressees, either constrained by experimental design or free to behave as they like (Bavelas et al., 2008). However, an alternative explaination is that the effect of visibility on gesture production is easier to find when speakers discuss spatial rather than abstract topics. To establish how these competing factors influence gesture production,we aimed to compare the effects of gesture visibility, addressee responsiveness, and the topic of discussion.

To answer the second research question, we have conducted a series of experiments manipulating both the visibility of the speaker's gesture, the visibility of the speaker and addressee's faces, and the responsive behaviour of the addressee. Furthermore, the experiments were conducted both with abstract topics being discussed (where gesture is not very suited to communicate information), and with spatial dynamic topics being described (where gesture is suited to communicate information. In both Chapters 3 and 4, speakers and described stimuli and produced gestures to an addressee. In Chapter 3, the experimenter acted as the addressee. The experimenter manipulated whether they were responsive or non-responsive. In Chapter 4, a second participant took on the role of the addressee. The second participant was either instructed to be conversational or silent (similar to responsive and non-responsive). In Chapter 3, speakers described abstract stimuli (Experiment 1) and spatial stimuli (Experiments 2 and 3). In Chapter 4, speakers only described spatial stimuli.

In Chapters 3 and 4, the effect of placing a screen between the speaker and addressee on the speaker's gesture production can be explained by the speakers' gestures being visible to the addressee (Chapter 4), and the addressee's face being visible to the speaker (Chapters 3 and 4), not whether the addressee could provide responses to the speaker. There is no evidence to suggest that the effect of visibility was due to the confounding of visibility and addressee responsiveness. First, the effect of visibility was found in both the non-responsive (silent) and responsive (conversational) addressee responsiveness conditions. There is no evidence that the effect of visibility on representational gesture only occurs when addressees are not responsive. Speakers produced more gestures when the addressee was visible than not visible even when addressees were free to be responsive in the conversational condition of Chapter 4. Second, on both Chapters 3 and 4, there was no effect of the addressee's responsiveness on the speaker's representational gesture rate. There was no evidence to suggest that responses from the addressee cause the speaker to produce more gestures. Together, these findings indicate that responses from the addressee causing the speaker to produce more gestures is not a plausible explanation for the effect of visibility on representational gestures.

We speculate that eye contact may be one of the factors contributing to a cue that triggers speakers to produce more gestures for the addressee. In the shoulder height screen conditions in Chapters 3 and 4, the speaker and the addressee could see each other's faces. In such situation, people tend to fixate on the addressee's eyes most of the time (Gullberg & Holmqvist, 2006). Previous research suggests eye contact has an important role in the effect of visibility on gesture. Speakers produce more gestures when they were able to make eye contact with the addressee than when the speaker and addressee could see each other's faces but not make eye contact (Mol et al., 2011). Mol et al. (2011) compared the rate at which speakers produced gestures when talking over webcams or using an "eye-catcher device" so that when the speaker and addressee are looking at each other on computer screens, it appears to the speaker and the addressee as if they are making eye contact. Speakers gestured more in the eye-catcher condition than the webcam condition. This suggests that eye contact contributes to a trigger for the heuristic. More broadly, we speculate that the face visibility cue concerns factors associated with the addressee's visual attention on the speaker.

In Chapter 3, the findings suggest that speakers consider gestures to be more valuable for communication when conveying dynamic spatial information than when conveying abstract information. Speakers only adapted their communication in response to the addressee's face when conveying dynamic spatial information. If speakers consider gesture to be valuable for conveying spatial information, then we can also partially explain why speakers produce more gestures when conveying dynamic spatial information rather than abstract information (Feyereisen & Havard, 1999). This increase in gesture rate for dynamic spatial information has been attributed to the increase in motor imagery elicited when discussing dynamic spatial stimuli (e.g. Hostetter & Alibali, 2008). As far as we know, there is no previous evidence to suggest that speakers produce more gestures when conveying dynamic spatial information *for communicative reasons*. Further research should directly test the claim that speakers consider gestures to be more valuable for communication when conveying dynamic spatial information than when conveying abstract information.

Overall, the findings suggest that several different factors affect co-speech gestures, with some factors interacting with each other. For example, the effect of visibility on gesture depends on the topic of discussion. The effect of visibility on gesture was observed when speakers were describing dynamic spatial stimuli but was not observed when speakers discussed abstract stimuli. Furthermore, the findings suggest that the effect of visibility is not dependent on whether the addressee is responsive or not.

6.1.3 The third research question

The third research question is how audience design behaviours make communication more effective. More specifically, it asked how placing gestures in the foreground of an interaction can help the gestures convey information to the addressee.

Speakers can place gestures in the “foreground” of the interaction to attract the attract the addressee’s attention and convey important information more clearly (Cooperrider, 2018; Streeck, 1993). Speakers can place gestures in the foreground by making gestures more visually prominent, by fixating their gaze on the gesture, or by referring to the gesture in speech using a demonstrative such as “like this” (Cooperrider, 2018).

To answer the third research question, we have conducted three experiments to investigate if placing gestures in the foreground conveys information to the addressee more effectively than background gestures. Participants watched videos of the experimenter describe cartoons either conveying target information via foregrounded or background gestures, or not conveying the target information. By comparing addressee comprehension in the foreground and background gesture conditions to the absent gesture condition, we tested if the gesture conveyed information to the addressees. By comparing addressee comprehension in the foreground condition to the background condition, we tested if gestures in the foreground conveyed information to the addressee more effectively than gestures in the background.

In Chapter 5, we found that it was *not* the case that foreground gestures communicated information more effectively than background gestures. The proportion of target information recalled in the foreground gesture conditions did not significantly differ from the proportion of target information recalled in the background gesture condition. For visually prominent gestures and gestures indicated by demonstratives, participants included the target information more often when the actor produced foreground gestures than when the actor did not produce any gestures. Furthermore, participants did not include the target information significantly more often when the actor produced background gestures than when the actor did not produce any gestures. Thus, foreground gestures communicate information to the addressee, while the background gestures did not significantly communicate information to the addressee. We found no effect of gaze fixation behaviours on addressee comprehension. Addressees did not include the target information more often when the experimenter produced gestures than when the experimenter did not produce gestures. Taken together, only the findings in Experiments 1 and 3 of Chapter 5 partially supports the claim that placing gestures in the foreground of the interaction benefits the addressee’s comprehension of the speaker’s message.

In Chapter 5, The descriptive trends in the data for visually prominent gestures and gestures indicated by demonstratives (but not gaze fixated gestures) support the claim that placing gestures in the foreground of the interaction benefits the addressee's comprehension of the speaker's message. For both visually prominent gestures and gestures indicated by demonstratives, the trend in the data suggests that foreground gestures lead to higher rates of target information being included in participants drawings than gestures in the background. Furthermore, the trend suggests that both foreground and background gestures can convey information to the addressee (more than in the absent gesture condition). Taken together, gestures which are more visually prominent or accompanied by a demonstrative are more effective at communicating information than gestures in the background of the interaction.

Overall, the findings from Chapter 5 did not provide evidence for the claim that foregrounded gestures communicate information more effectively than background gestures. There is no evidence in Chapter 5 to suggest that speakers may design their gestures to convey important information more clearly (Cooperrider, 2018; Streeck, 1993). While placing gestures in the foreground of the interaction may provide a clear indicator that the speaker produces the gesture to communicate information to the addressee (Streeck, 1993), the current study failed to find unequivocal evidence that addressee's benefitted from speakers placing gestures in the foreground.

6.2 The findings in relation to the literature

6.2.1 Implications for the audience design literature

The findings in the current thesis have implications for the theory of audience design behaviours. Chapters 3 and 4 are the first to find empirical evidence for a cue-based heuristic audience design behaviour. The findings support the claim that Speakers can take advantage of cues available in their interaction with the addressee (Shintel & Keysar, 2009). Such cues can trigger speakers to change their behaviour in pre-determined ways. Speakers do not have to use these cues to first update a model of the addressee (as in the co-presence cue, discussed above) and then adjust communicative behaviour based on the model. Cues that directly trigger audience design behaviours might include addressee feedback, visual behaviours such as eye gaze and gesture, and the addressee's status as a child or foreign speaker.

The addressee's face as a cue fits with speaker's focus on the addressee's face in prior research. Addressees use facial expressions to provide feedback to the speaker, which can impact the speaker's communication (Bavelas et al., 2000; Bavelas & Gerwing, 2011) and speakers are sensitive to the addressee's gaze in conversation (C. Goodwin, 1986; M. H.

Goodwin & Goodwin, 1986). As speakers are paying attention to the face to coordinate verbal communication any way, speakers can monitor the addressee's face for cues without any additional cognitive cost. Furthermore, speakers can check if the addressee is paying visual attention to the speaker, a pre-requisite for visual communication. Finally, in most everyday situations, when the addressee's face can be seen, the addressee can see the speaker's gestures. Visibility of the addressee's face is an ideal cue; it is low cost, it checks a key pre-requisite for visual communication, and it works in most cases.

The clear empirical evidence for cue-based audience design in Chapters 3 and 4 provide additional support for cue-based audience designs in other aspects of communication. Speakers can attenuate both speech and gesture when retelling a story to the same addressee a second time by using the identity of the addressee as a cue (Galati & Brennan, 2010, 2014). Speakers can orient gesture production towards the addressee by using the location of the addressee as a cue (Özyürek, 2002). Speakers could use the identity of the addressee as a cue to trigger infant directed speech (Snow, 1972). The Lombard effect can also be explained by the cue-based heuristic view. Speakers could use the presence of the addressee as a cue increase the modulations made in response to background noise. If the cues necessary to trigger a behaviour are present, then it is plausible that an audience design behaviour can be explained by a cue-based heuristic mechanism. The previous literature argued for such cue-based heuristics purely on the theoretical basis that perspective taking would be too computationally costly, and thus computationally simpler heuristics must be at play. The current study for the first time provided direct empirical evidence for a cue-based heuristic for gesture production, which makes it more plausible that cue-based heuristics are also at play in other aspects of communication.

Furthermore, the findings in Chapter 4 are the first to provide evidence that speakers can both respond to cues from the addressee through heuristic processes and take the perspective of the addressee to communicate more effectively with the addressee. Previous findings have been compatible with a perspective taking explanation (e.g. Alibali et al., 2001), but the findings could also be explained by a cue-based heuristic explanation. The only previous comparison of a perspective taking and heuristic based explanation found no evidence that speakers took the perspective of the addressee (Barker & Kita, 2021). Chapter 4 builds upon the findings of Chapter 3 by suggesting that speakers can both respond to cues and take the perspective of the addressee to adjust the same communicative behaviour.

Cue-based heuristics can be incorporated into a mechanistic framework for audience design (Ferreira, 2019). While Ferreira's framework is concerned with language audience design, it is possible to incorporate non-verbal behaviours into the same model. One source of audience design in the model comes from pre-learned automatic encoding strategies. Speakers can apply learned behaviours to create rules in a feedforward language production process. These rules establish the connection between a cue and a specific behaviour to form a heuristic mechanism. We suggest that the cue-based heuristic demonstrated in the current study fits into the mechanistic framework for audience design as a pre-learned automatic encoding strategy. It is plausible that other heuristic strategies also result from learned rules triggering automatic encoding strategies. For example, speakers attenuate their gestures when repeating a story to an addressee who has heard the story from the speaker before (Galati & Brennan, 2014). The speakers may have already created rules from prior experience about attenuating gesture when retelling information. If this is the case, speakers would not need to consult a model of the addressee and infer that gesture can be attenuated. Instead, the status of the addressee as knowledgeable about the information triggers the pre-learned strategy.

6.2.2 Implications for the gesture production literature

Three of the key findings in the current thesis have implications for theories of gesture production in communicative settings. First, the findings from Chapters 3 and 4 support the claims that visibility of gestures and visibility of the addressee's face can affect gesture production. Second, the findings from Chapter 3 suggest that speakers consider gestures to be more valuable for communication when conveying dynamic spatial information than when conveying abstract information. Third, the findings of Chapters 3 and 4 do not replicate the finding of previous studies that speakers produce more gestures when conversing with a responsive addressee than a non-responsive addressee.

The cue-based heuristic view can explain the previous findings of the effect of visibility on gesture. Previous research into the effect of visibility on gesture production has always confounded gesture visibility with facial visibility, (e.g. Alibali et al., 2001; Mol et al., 2009). When the addressee could see the speaker's gestures, the speaker could also see the addressee's face. According to the results of the current study, facial visibility cues the speaker to produce more gestures. Additionally, previous research with speakers discussing abstract topics did not produce more gestures when the addressee could see the speaker (e.g. Bavelas et al., 1992; Rimé, 1982) which can be explained by speakers not finding gestures as useful for communicating abstract topics. Therefore, the heuristic based on facial visibility and spatial content parsimoniously explain the results from the current study and previous studies on the effect of visibility on gesture.

The current study appears to contradict prior findings suggesting speakers are sensitive to whether the speaker's gestures can be seen (Mol et al., 2011). Over a webcam, speakers produced more gestures when the addressee could see the speaker. Speakers did not produce more gestures when the speaker could see the addressee. The finding appears to suggest that speakers are sensitive to whether their own gestures are useful for communication, and do not respond to the addressee's face as a cue. However, the findings of the previous study can be explained by the use of a web cam and the experimental instructions. The addressee's face may not have acted as a cue because eye contact was not possible. When eye contact could be made, speakers produced far more gestures when the speaker knew their gestures could be seen by the addressee (Mol et al., 2011). Furthermore, speakers may have altered their gesture production after the experimenter highlighted whether the addressee could see the speaker or not by displaying the addressee's point of view. It is plausible that showing the addressee's perspective reduced the cognitive cost of perspective taking. Therefore, speakers may have produced more gestures in response to being shown by the experimenter that they were visible, or supressed gesture production when shown that they were not visible to the addressee.

The findings in Chapters 3 and 4 suggest that the topic of discussion, and not the behaviour of the addressee can explain the why some studies fail to find an effect of visibility on representational gesture rate. Studies that have not found an effect of visibility have used static spatial or abstract stimuli to elicit gesture (e.g. Bavelas et al., 2008), while studies that have found an effect of visibility have used dynamic spatial and motoric stimuli to elicit gesture (e.g. Alibali et al., 2001), If speakers produce more gestures to communicate information when discussing dynamic spatial and motoric topics, then the effect of visibility on gesture will be stronger. Gestures conveying spatial or motoric information result in gesture becoming more useful for communication (Hostetter, 2011), and speakers produce gestures more when describing spatial information than non-spatial information (Alibali, 2005; Rauscher et al., 1996). It is plausible that previous free-dialogue studies failed to find an effect of visibility on representational gesture because speakers were not motivated to produce gestures to benefit communication when describing static spatial or abstract topics. The current study may have found an effect of visibility in a free-dialogue conversational condition because the speakers were motivated to produce gestures to benefit communication when describing motoric and dynamic spatial topics.

Neither Chapter 3 or 4 replicated previous findings that speakers produce more gestures when conversing with a responsive addressee than a non-responsive addressee (e.g. Bavelas & Gerwing, 2011; Beattie & Aboudan, 1994; Jacobs & Garnham, 2007). Speaker's gesture production was unaffected by the addressee's responsive behaviour. If addressees are able to contribute information to the interchange with the speaker as in previous studies (Bavelas & Gerwing, 2011; Beattie & Aboudan, 1994; Jacobs & Garnham, 2007) then responsive addressee's may trigger speakers to produce more gestures.

The findings of Chapters 3 and 4 also have implications for whether speakers intend to communicate information for an addressee or not. In some situations, researchers have ascribed intentionality or a lack of it to specific behavioural acts. For example, in children, protoimperative pointing (pointing to something to obtain that something) is not seen as an intentional act of communication in young infants as it does not require any knowledge of a mental agent (Tomasello et al., 2007). In comparison, protodeclarative pointing (pointing to share attention with another person) is considered intentionally communicative because it does require knowledge of the other person as a mental agent (Tomasello et al., 2007). However, it can be more difficult to ascribe intentionality to behaviours which may occur automatically. We have argued that heuristics are low cost, produced automatically in response to specific stimuli, and can be overridden by a strong motivation to do so. Automatic processes are also thought to occur without intention. However, it is possible to distinguish intention, awareness, efficient and control as separate issues (Bargh, 1989). It is argued that just because a process is automatic, it does not follow that the process is intentional or now. It is outside of the findings of the current studies to ascribe intentionality (or not) to the speaker's communicative behaviours throughout the different experiments.

6.2.3 Implications for the gesture comprehension literature

The findings Chapter 5 have implications for the gesture comprehension literature. Chapter 5 is the first study to investigate the effect of foregrounding gestures on addressee comprehension. Previous studies have used foregrounding behaviours as indicators that speakers intend to communicate information. The use of foreground behaviours as measures of communicative intent assumes that foregrounding behaviours are designed to help convey information to the addressee.

Chapter 5 is the first to investigate the effect of gesture visual prominence on addressee's comprehension. Previous studies have assumed that larger gestures are an indicator that the speaker intends to communicate informaiton to the addressee (e.g. Bavelas et al., 2008; Holler & Wilkin, 2011). The findings of Chapter 5 did not provide evidence for the claim that more visually prominent gestures communicate more effectively than non-visually prominent gestures, although only visually prominent gestures significantly conveyed information to the addressee. Further work is needed to establish the role visually prominent gestures play in communicating information and signaling the speaker's intention to communicate.

Chapter 5 is also the first to investigate the effect of producing demonstratives alongside gestures on addressee's comprehension. Previous studies have assumed that gestures which are accompanied by demonstratives are an indicator that the speaker intends to communicate informaiton to the addressee using gesture (Bavelas et al., 2008; de Ruiter et al., 2012; Emmorey & Casey, 2001). The findings of Chapter 5 did not provide evidence for the claim that gestures which are accompanied by demonstratives communicate information more effectively than gestures which are not accompanied by demonstratives. As the trend in the data suggests that gestures accompanied by demonstratives do benefit communication more than gestures which are not accompanied by demonstratives, further research is needed to establish how demonstratives benefit communication.

Chapter 5 adds to the previous research on the effect of speakers fixating their gaze on their gestures. Speakers can fixate their gaze on their own gestures to draw the addressee's attention to gesture (Gullberg & Kita, 2009). However, it was unclear if the addressees benefitted from speakers fixating on the gestures, as gesture duration was confounded with the effect of gaze fixation (Gullberg & Kita, 2009). Furthermore, it was unclear if the effect of gaze fixation on addressee comprehension generalised beyond directional information. The findings did not provide evidence for the claim that the effect of gaze fixation on addressee comprehension generalised beyond directional information.

6.3 **Further research**

Further research should continue to examine the competing factors which affect gesture production. Chapters 3 and 4 leave the relationships unresolved between the topic of discussion, the responses provided by the addressee, and speakers production of gestures.

Further research should directly test if speakers produce more gestures when discussing spatial information because gesture is particularly suited to communicate information. The finding that speakers produce more gestures when discussing spatial information has been attributed to the spatial information eliciting gestures (e.g. Hostetter & Alibali, 2008). However, the increased use of gestures when discussing spatial information could partically be explained by how suitable gesture is for communicating information (Hostetter, 2011).

Further research should investigate why the manipulations of addressee responses failed to elicit an effect on the speaker's gesture production. Both Chapters 3 and 4 failed to replicate previous findings that speakers produce more gestures when conversing with a responsive addressee than a non-responsive addressee (e.g. Bavelas & Gerwing, 2011; Beattie & Aboudan, 1994; Jacobs & Garnham, 2007). Our explanation that addressee responses only affect speaker's gesture production when addresses can act as a contributor of information to the discourse needs to be directly tested. This finding would be able to reconcile the failure to find an effect of addressee responsiveness on gesture production in Chapters 3 and 4 with the prior literature on addressee responsiveness.

Further research could more closely examine the addressee's behaviour in chapters 3 and 4. The findings may have been influenced by differences in the addressee's behaviour between the different conditions. It is possible that the addressee may have used more of one type of verbal response than another. Furthermore, it is possible that the addressee may have used different visual responses which affect gesture production differently. Such differences will not have been captured by our control measure of addressee responsiveness rating. However, any such differences would have produced much smaller effects than the addressee responsiveness manipulation. If small variations in addressee responsiveness had a significant effect speaker's gesture production, then larger variations in addressee responsiveness should have had an effect on addressee responsiveness. As no effect of addressee responsiveness was found, it suggests that speaker's gesture production was not subject to variation in the addressee's responsiveness. Further research could more closely examine the behaviour of the addressee to look for discrepancies in the addressee's behaviour.

Further research using Chapter 3 and Chapter 4s' data set could examine the content of the gestures to establish if the information expressed in gesture changes in response to cues from the addressee. It's plausible that the content of gesture changes when speakers are communicating information to the addressee, especially as more informative gestures benefit communication (Krason et al., 2021). It is possible that while speakers gestured more when the addressee's face was visible to the speaker, the gestures may have only been more informative to the addressee when the speaker's gestures were visible. Further research could establish if the speaker's gestures are more informative when the gestures are visible to the addressee, or if the addressee's face is visible to act as cue to the speaker.

Speakers may not have produced different rates of gesture in the responsive and non-responsive conditions because the addressee's behaviour matched the speakers' expectations in both conditions. The manipulation of addressee responsiveness differed from previous studies' manipulations. For example, Bavelas et al. (2008) did not have a non-responsive addressee in the equivalent non-responsiveness condition. The speaker talked to a tape recorder, instead of a non-responsive addressee. While it is possible that speakers produce more gestures the more responsive an addressee is, when no responses are provided, the speaker may change their behaviour to adapt to the complete lack of feedback. For example speaker's with distracted addressees perform worse on narrative tasks (Bavelas et al., 2000). In previous studies with silent addressees however, speakers produced a higher rate of gestures when talking to responsive addressees than non-responsive addressees (Beattie & Aboudan, 1994). The finding suggests that speakers still produce fewer gestures in response to silent addressees than responsive addressees. The findings of the current study may differ from previous studies because speakers expected responsive or non-responsive behaviours from the addressee and adapted to them. Speakers gesture more frequently when their expectations were consistent with addressees' behaviour (Kuhlen et al., 2012). For example, when they expected distracted addressees and addressees were indeed distracted. In the current study, speakers may have expected the addressee to be responsive or not after the practice session. This finding can explain why no effect of responsiveness was found; speakers adapted to the addressee's lack of responsiveness based off of their expectations.

Further research should examine the effect of motivation of speakers' use of the heuristic and perspective taking mechanisms. It is plausible that speakers who are motivated to communicate effectively are more likely to take the perspective of the addressee than unmotivated speakers. As previous suggested, if speakers believed it was important to clearly convey the information to the addressee, then speakers may have been more motivated to take the perspective of the addressee to ensure effective communication. Furthermore, it is also plausible that speakers were more motivated to take the perspective of the addressee when perspective taking is less cognitively demanding because the speaker and addressee are peers. Unfortunately, there is limited evidence to directly support either factor that may have led participants to take the perspective of the addressee more in Chapter 4 than in the experiments in Chapter 3. Further research should investigate the effects of motivation on perspective taking.

Further research should investigate the possibility that eye contact rather than facial visibility in general acts as the cue for visual communication heuristics. We speculated that eye contact may be one of the factors contributing to a cue that triggers speakers to produce more gestures for the addressee. In the shoulder height screen conditions in Chapters 3 and 4, the speaker and the addressee could see each other's faces. In such situation, people tend to fixate on the addressee's eyes most of the time (Gullberg & Holmqvist, 2006). Previous research suggests eye contact has an important role in the effect of visibility on gesture. Speakers produce more gestures when they were able to make eye contact with the addressee than when the speaker and addressee could see each other's faces but not make eye contact (Mol et al., 2011). Mol et al. (2011) compared the rate at which speakers produced gestures when talking over webcams or using an "eye-catcher device" so that when the speaker and addressee are looking at each other on computer screens, it appears to the speaker and the addressee as if they are making eye contact. Speakers gestured more in the eye-catcher condition than the webcam condition. This suggests that eye contact contributes to a trigger for the heuristic. More broadly, we speculate that the face visibility cue concerns factors associated with the addressee's visual attention on the speaker. Further research should directly test this speculation.

Further research should more closely examine the effect of gestural audience design behaviours on addressee comprehension. The findings in Chapter 5 only partially suggest that placing gestures in the foreground of the interaction helps the gestures convey information to the addressee. these findings did not provide evidence for the idea that speakers placed their gestures in the foreground for communicative purposes (Cooperrider, 2018). The power analysis used to predict the required number of participants to detect the effect was based on a study with similar method except that participants were in the presence of the experimenter while taking part rather than taking part online (Theron-Grimaldi, 2021). The very different levels of target information being included (50% vs 28% highest) suggest that speakers were less motivated to attend to the stimuli in Chapter 5. This methodological difference could have greatly diluted the effect of foregrounding gestures on addressee comprehension. Therefore, further research should be conducted where addressees are more motivated to attend to the stimuli.

6.4 Conclusion

The current thesis investigated how speakers design their gestures to communicate more effectively with an addressee. We found that speakers can use two mechanisms to produce audience design behaviours. The first is a heuristic process that triggers in the presence of addressee related cues. The second is a more cognitively intensive mechanisms where the speaker is motivated to take the perspective of the addressee and infer how to design communicative behaviours. We also found that speakers adjust their gesture production to communicate more effectively depending on several factors including mutual visibility of gesture, mutual visibility of faces, and the topic of discussion. Furthermore, we found that visually prominent gestures and gestures indicated by a demonstrative can convey information to the addressee, but that the effect of foreground gestures on addressee comprehension needs to be further examined. Taken together, the findings suggest that speakers adjust their gestures to communicate with the addressee, producing more gestures when the gestures can serve a communicative function. However, the effect of the gestural audience design behaviours on addressee comprehension should be further examined.

Bibliography and References

Alibali, M. W. (2005). Gesture in Spatial Cognition: Expressing, Communicating, and Thinking About Spatial Information. *Spatial Cognition and Computation*, *5*(4), 307–331. https://doi.org/10.1207/s15427633scc0504

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen. *Journal of Memory and Language*, *44*(2), 169–188. https://doi.org/10.1006/jmla.2000.2752

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. https://doi.org/10.1080/016909600750040571

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in Mathematics Teaching and Learning: Evidence From Learners' and Teachers' Gestures. *Journal of the Learning Sciences*, *21*(2), 247–286. https://doi.org/10.1080/10508406.2011.611446

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138–1144. https://doi.org/10.1177/0956797611417722

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415–419. https://doi.org/10.1111/j.0956-7976.2004.00694.x

Bangerter, A., & Chevalley, E. (2007). Pointing and describing in referential communication: When are pointing gestures used to communicate? *MOG 2007 Workshop on Multimodal Output Generation*, 17–28. http://let.uvt.nl/research/ti/iso-tdg3/

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition.tle. *Unintended Thought*, 3–51.

Barker, J., & Kita, S. (2021). The Addressee's Face as a Trigger for Gesture Production: Evidence for Cue-Based Heuristics in Audience Design. *Manuscript Submitted for Publication*.

Barr, D. J., & Keysar, B. (2006). Perspective Taking and the Coordination of Meaning in Language Use. In *Handbook of Psycholinguistics* (pp. 901–938). Academic Press. https://doi.org/10.1016/B978-012369374-7/50024-9

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

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bavelas, J. B. (2007). Face-to-face dialogue as a micro-social context. In S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the Dynamic Dimension fr or Language Essays in honor of David McNeill* (pp. 127–146). http://web.uvic.ca/psyc/bavelas/2007Microsocial.pdf

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. In *Personality and Social Psychology Bulletin* (Vol. 21, Issue 4, pp. 394–405). https://doi.org/10.1177/0146167295214010

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*(4), 469–489. https://doi.org/10.1080/01638539209544823

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952. https://doi.org/10.1037/0022-3514.79.6.941

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a communication process: The role of gaze. *Journal of Communication*, *52*(3), 566–580. https://doi.org/10.1111/j.1460-2466.2002.tb02562.x

Bavelas, J. B., & Gerwing, J. (2011). The listener as addressee in face-to-face dialogue. *International Journal of Listening*, *25*(3), 178–198. https://doi.org/10.1080/10904018.2010.508675

Bavelas, J. B., Gerwing, J., Allison, M., & Sutton, C. (2011). *Dyadic evidence for grounding with abstract deictic gestures*. http://web.uvic.ca/psyc/bavelas/2011Floor_plan.pdf

Bavelas, J. B., Gerwing, J., & Healing, S. (2014a). Hand Gestures and Facial Displays in Conversational Interaction. In T. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology* (pp. 111–130). Oxford University Press.

https://doi.org/10.1093/oxfordhb/9780199838639.013.008

Bavelas, J. B., Gerwing, J., & Healing, S. (2014b). Hand and Facial Gestures in Conversational Interaction. In T. M. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.013.008

Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520. https://doi.org/10.1016/j.jml.2007.02.004

Bavelas, J. B., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, *13*(1), 63–92. https://doi.org/10.1075/gest.13.1.03bav

Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, *99*(3–4), 239–272. https://doi.org/10.1515/semi.1994.99.3-4.239

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, *123*(1–2), 1–30. https://doi.org/10.1515/semi.1999.123.1-2.1

Beattie, G., & Shovelton, H. (2005). Why the spontaneous images created by the hands during talk can help make TV advertisements more effective. In *British Journal of Psychology* (Vol. 96, Issue 1, pp. 21–37). John Wiley & Sons, Ltd. https://doi.org/10.1348/000712605x103500

Blokpoel, M., van Kesteren, M., Stolk, A., Haselager, P., Toni, I., & van Rooij, I. (2012). Recipient design in human communication: simple heuristics or perspective taking? *Frontiers in Human Neuroscience*, *6*, 253. https://doi.org/10.3389/fnhum.2012.00253

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for “motionese”: Modifications in mothers’ infant-directed action. *Developmental Science*, *5*(1), 72–83. https://doi.org/10.1111/1467-7687.00211

Breckinridge Church, R., Garber, P., & Rogalski, K. (2007). The role of gesture in memory and social communication. *Gesture*, *7*(2), 137–158. https://doi.org/10.1075/gest.7.2.02bre

Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in*

Cognitive Science, *1*(2), 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making Children Gesture Brings Out Implicit Knowledge and Leads to Learning. *Journal of Experimental Psychology: General*, *136*(4), 539–550. https://doi.org/10.1037/0096-3445.136.4.539

Broaders, S. C., & Goldin-Meadow, S. (2010). Truth is at hand: How gesture adds information during investigative interviews. *Psychological Science*, *21*(5), 623–628. https://doi.org/10.1177/0956797610366082

Brunner, L. J. (1979). Smiles can be back channels. *Journal of Personality and Social Psychology*, *37*(5), 728–734. https://doi.org/10.1037/0022-3514.37.5.728

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, *96*(1), 168–174.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*(4), 706–723. https://doi.org/10.1037/a0013157

Chu, M., & Kita, S. (2011). The Nature of Gestures' Beneficial Role in Spatial Problem Solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(2), 257–270. https://doi.org/10.1037/xlm0000168

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*(2), 694–709. https://doi.org/10.1037/a0033861

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43–71. https://doi.org/10.1016/0010-0277(86)90053-3

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*(3), 324–347. https://doi.org/10.1037/rev0000026

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K.

Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge University Press.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*(C), 287–299. https://doi.org/10.1016/S0166-4115(09)60059-5

Cohen, A. A. (1977). The Communicative Functions of Hand I1lustrators. *Journal of Communication*, *27*(4), 54–63. https://doi.org/10.1111/j.1460-2466.1977.tb01856.x

Cohen, A. A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, *28*(2), 276–279. https://doi.org/10.1037/h0035792

Cooperrider, K. (2018). Foreground gesture, background gesture. *Gesture*, *16*(2), 176–202. https://doi.org/10.1075/gest.16.2.02coo

Craycraft, N. N., & Brown-Schmidt, S. (2018). Compensating for an Inattentive Audience. *Cognitive Science*, *42*(5), 1504–1528. https://doi.org/10.1111/cogs.12614

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin*, *145*(8), 765–784. https://doi.org/10.1037/bul0000202

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (Vol. 50, Issue 1, pp. 284–311).

de Ruiter, J. P. (2006). Can gesticulation help aphasic people speak, or rather, communicate? In *Advances in Speech Language Pathology* (Vol. 8, Issue 2, pp. 124–127). https://doi.org/10.1080/14417040600667285

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

Deliens, G., Antoniou, K., Clin, E., & Kissine, M. (2017). Perspective-taking and frugal strategies: Evidence from sarcasm detection. *Journal of Pragmatics*, *119*, 33–45. https://doi.org/10.1016/j.pragma.2017.08.002

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283–292.

https://doi.org/10.1037/h0033031

Ekman, P., & Friesen, W. V. (1969). The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica*, *1*(1), 49–98. https://doi.org/10.1515/semi.1969.1.1.49

Emmorey, K., & Casey, S. (2001). Gesture, thought and spatial language. *Gesture*, *1*(1), 35–50. https://doi.org/10.1075/gest.1.1.04emm

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge University Press. https://doi.org/10.1017/CBO9780511576737

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

Ferreira, V. S. (2019). A Mechanistic Framework for Explaining Audience Design in Language Production. *Annual Review of Psychology*, *70*(1), 29–51. https://doi.org/10.1146/annurev-psych-122216-011653

Feyereisen, P., & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior*, *23*(2), 153–171. https://doi.org/10.1023/A:1021487510204

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 319–325. https://doi.org/10.1002/wcs.1226

Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, *19*(6), 509–525. https://doi.org/10.1002/ejsp.2420190603

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35–51. https://doi.org/10.1016/j.jml.2009.09.002

Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, *29*(4), 435–451. https://doi.org/10.1080/01690965.2013.796397

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the lombard effect. *Journal of Speech, Language, and*

Hearing Research, *53*(3), 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138)

Gerrig, R. J., & Littman, M. L. (1990). Disambiguation by community membership. *Memory & Cognition*, *18*(4), 331–338. https://doi.org/10.3758/BF03197122

Gerwing, J., & Bavelas, J. B. (2004). Linguistic influences on gesture's form. *Gesture*, *4*(2), 157–195. https://doi.org/10.1075/gest.4.2.04ger

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429. https://doi.org/10.1016/S1364-6613(99)01397-2

Goldin-Meadow, S., Nusbaum, H. C., Kelly, S. D., & Wagner, S. (2001). Explaining Math: Gesturing Lightens the Load. *Psychological Science*, *12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Goodwin, C. (1986). Gestures as a resource for the organization of mutual orientation. *Semiotica*, *62*(1–2), 29–50. https://doi.org/10.1515/semi.1986.62.1-2.29

Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, *62*(1–2), 51–76. https://doi.org/10.1515/semi.1986.62.1-2.51

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology*, *5*(2), 189–195. https://doi.org/10.1002/ejsp.2420050204

Grice, H. P. (1975). Logic and Conversation. , Speech Acts. *In P. Cole, & J. L. Morgan. (Eds.), Syntax and Semantics, New York: Academic Press.*, *Vol. 3*, 41–58. https://brill.com/view/book/edcoll/9789004368811/BP000003.xml

Guérin, V. (2015). Demonstrative verbs: A typology of verbal manner deixis. *Linguistic Typology*, *19*(2), 141–199. https://doi.org/10.1515/lingty-2015-0006

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, *14*(1), 53–82. https://doi.org/10.1075/pc.14.1.05gul

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, *33*(4), 251–277. https://doi.org/10.1007/s10919-009-0073-2

Hilliard, C., O'Neal, E., Plumert, J., & Cook, S. W. (2015). Mothers modulate their gesture independently of their speech. *Cognition*, *140*, 89–94. https://doi.org/10.1016/j.cognition.2015.04.003

Hoetjes, M., Krahmer, E., & Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, *57*, 257–267. https://doi.org/10.1016/j.specom.2013.06.007

Hoetjes, M., & Masson-Carro, I. (2017). Under load: The effect of verbal and motoric cognitive load on gesture production. *Journal of Multimodal Communication Studies*, *4*(1–2), 29–36.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*(1), 4–27. https://doi.org/10.1177/0261927X06296428

Holler, J., Tutton, M., & Wilkin, K. (2011). Co-speech gestures in the process of meaning coordination. *2nd GESPIN-Gesture & Speech in Interaction*. http://pubman.mpdl.mpg.de/pubman/item/escidoc:1105560/component/escidoc:1105561/Holler, Tutton & Wilkin 2011_gespin-1.pdf

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, *43*(14), 3522–3536. https://doi.org/10.1016/j.pragma.2011.08.002

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Hostetter, A. B. (2011). When Do Gestures Communicate? A Meta-Analysis. *Psychological Bulletin*, *137*(2), 297–315. https://doi.org/10.1037/a0022128

Hostetter, A. B. (2014). Action attenuates the effect of visibility on gesture rates. *Cognitive Science*, *38*(7), 1468–1481. https://doi.org/10.1111/cogs.12113

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, *7*(1), 73–95. https://doi.org/10.1075/gest.7.1.05hos

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. In *Psychonomic Bulletin and Review* (Vol. 15, Issue 3, pp. 495–514). Springer-Verlag. https://doi.org/10.3758/PBR.15.3.495

Hostetter, A. B., & Alibali, M. W. (2011). Cognitive skills and gesture–speech redundancy: Formulation difficulty or communicative strategy? *Gesture*, *11*(1), 40–60. https://doi.org/10.1075/gest.11.1.03hos

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. In *Psychonomic Bulletin and Review* (Vol. 26, Issue 3, pp. 721–752). Springer New York LLC. https://doi.org/10.3758/s13423-018-1548-0

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). Does Sitting on Your Hands Make You Bite Your Tongue ? The Effects of Gesture Prohibition on Speech During Motor Descriptions. *Proceedings of the Annual Meeting of the Cognitive Science*, *29*, 29. https://cloudfront.escholarship.org/dist/prd/content/qt6g02f7bv/qt6g02f7bv.pdf

Hostetter, A. B., Alibali, M. W., & Schrager, S. M. (2011). If you don't already know, I'm certainly not going to show you! In G. Stam & M. Ishino (Eds.), *Integrating Gestures: The interdisciplinary nature of gesture* (pp. 61–74). John Benjamins.

Hostetter, A. B., & Boncoddo, R. (2017). Gestures highlight perceptual–motor representations in thinking. In R. Breckinridge Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why Gesture?: How the hands function in speaking, thinking and communicating.* (pp. 155–174). PA: Benjamins. https://books.google.co.uk/books?hl=en&lr=&id=hYerDgAAQBAJ&oi=fnd&pg=PA155&dq=Hostetter,+A.+B.,+%2526+Boncoddo,+R.+(2017).+Gestures+highlight+perceptual–motor+representations+in+thinking.+In+R.+B.+Church,+M.+W.+Alibali,+%2526+S.+D.+Kelly+(Eds.),+Why+gesture%253

Hostetter, A. B., & Potthoff, A. L. (2012). Effects of personality and social situation on representational gesture production. *Gesture*, *12*(1), 62–83. https://doi.org/10.1075/gest.12.1.04hos

Howell, D. C. (2017). *Fundamental statistics for the behavioural sciences* (9th Editio). Cengage Learning.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, *56*(2), 291–303. https://doi.org/10.1016/j.jml.2006.07.011

Kelly, S. D., Byrne, K., & Holler, J. (2011). Raising the Ante of Communication: Evidence for

Enhanced Gesture Use in High Stakes Situations. *Information*, *2*(4), 579–593. https://doi.org/10.3390/info2040579

Kendon, A. (1994). Do Gestures Communicate?: A Review. *Research on Language and Social Interaction*, *27*(3), 175–200. https://doi.org/10.1207/s15327973rlsi2703_2

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

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K., & Green, A. (2009). Neural interaction of speech and gesture: Differential activations of metaphoric co-verbal gestures. *Neuropsychologia*, *47*(1), 169–179. https://doi.org/10.1016/j.neuropsychologia.2008.08.009

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162–185). Cambridge University Press. https://doi.org/10.1017/cbo9780511620850.011

Kita, S., Alibali, M. W., & Chu, M. (2017). How Do Gestures Influence Thinking and Speaking? The Gesture-for-Conceptualization Hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059.supp

Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes*, *24*(5), 761–775. https://doi.org/10.1080/01690960802327971

Kita, S., & Essegbey, J. (2001). Pointing left in Ghana: How a taboo on the use of the left hand influences gestural practice. *Gesture*, *1*(1), 73–95. https://doi.org/10.1075/gest.1.1.06kit

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

Kita, S., Van Gijn, I., & Hulst, H. van der. (1997). Movement phase in signs and co-speech gestures, and their transcriptions by human coders. *Gesture and Sign Language in Human-Computer Interaction. Bielefeld Gesture Workshop. September 17-19*, 23–35. https://pdfs.semanticscholar.org/d2ca/acf62920f45e110e8faa3369ea1187ad2402.pdf

Krason, A., Fenton, R., Varley, R., & Vigliocco, G. (2021). The role of iconic gestures and mouth movements in face-to-face communication. *Psychonomic Bulletin and Review*.

https://doi.org/10.3758/s13423-021-02009-5

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The Communicative Value of Conversational Hand Gesture. *Journal of Experimental Social Psychology*, *31*(6), 533–552. https://doi.org/10.1006/jesp.1995.1024

Krauss, R. M., & Fussell, S. R. (1991). Perspective-Taking in Communication: Representations of Others' Knowledge in Reference. In *Social Cognition* (Vol. 9, Issue 1). Guilford . https://doi.org/10.1521/soco.1991.9.1.2

Krauss, R. M., & Fussell, S. R. (2004). Constructing shared communicative environments. In L. Resnick, J. Levine, & S. Behrend (Eds.), *Perspectives on socially shared cognition.* (pp. 172–200). American Psychological Association. https://doi.org/10.1037/10096-008

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, *47*(7), 567–587. https://doi.org/10.1080/01638530903441339

Kuhlen, A. K., Galati, A., & Brennano, S. E. (2012). Gesturing integrates top-down and bottom-up information: Joint effects of speakers' expectations and addressees' feedback. *Language and Cognition*, *4*(1), 17–41. https://doi.org/10.1515/langcog-2012-0002

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* University of Chicago Press.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press.

Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales Des Maladies de L'Oreille et Du Larynx*, *37*, 101–119.

McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, *32*(7), 855–878. https://doi.org/10.1016/S0378-2166(99)00079-X

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*, *24*(2), 131–150. https://doi.org/10.1023/a:1006657929803

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* (Vol. 27, Issue 4). University of Chicago Press. https://doi.org/10.2307/1576015

McNeill, D. (2003). Pointing and Morality in Chicago. In S. Kita (Ed.), *Pointing* (pp. 301–314). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410607744-16

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press. https://doi.org/10.5860/choice.43-5099

McNeill, D., Cassell, J., & Levy, E. T. (1993). Abstract deixis. *Semiotica*, *95*(1–2), 5–20. https://doi.org/10.1515/semi.1993.95.1-2.5

McNeill, D., Cassell, J., & McCullough, K. E. (1994). Communicative Effects of Speech-Mismatched Gestures. *Research on Language and Social Interaction*, *27*(3), 223–237. https://doi.org/10.1207/s15327973rlsi2703_4

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, *22*(4), 473–500. https://doi.org/10.1080/01690960600696916

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*(2), 9–4. https://doi.org/10.1075/gest.4.2.02mel

Mol, L., & Kita, S. (2012). Gesture structure affects syntactic structure in speech. *34th Annual Conference of the Cognitive Science Society*, 761–766. http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CCYQFjAA&url=http://mindmodeling.org/cogsci2012/papers/0141/paper0141.pdf&ei=f3WJUJC2MI67hAfM1oCQBA&usg=AFQjCNETutOuYPmqtBAAE3eZlA7Q01jSOQ&sig2=tk_zME8jrvlmGAi6JmokX

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009a). Communicative Gestures and Memory Load. *Design*, 1569–1574.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009b). The communicative import of gestures: Evidence from a comparative analysis of human–human and human–machine interactions. *Gesture*, *9*(1), 97–126. https://doi.org/10.1075/gest.9.1.04mol

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2011). Seeing and being seen: The effects on gesture production. *Journal of Computer-Mediated Communication*, *17*(1), 77–100. https://doi.org/10.1111/j.1083-6101.2011.01558.x

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology*, *117*(3), 411–424. https://doi.org/10.2307/4149008

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nummenmaa, L., Smirnov, D., Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Sams, M., & Hari, R. (2014). Mental action simulation synchronizes action-observation circuits across individuals. *Journal of Neuroscience*, *34*(3), 748–757. https://doi.org/10.1523/JNEUROSCI.0352-13.2014

O'Carroll, S., Nicoladis, E., & Smithson, L. (2015). The effect of extroversion on communication: Evidence from an interlocutor visibility manipulation. *Speech Communication*, *69*, 1–8. https://doi.org/10.1016/j.specom.2015.01.005

Osterhout, L., & Swinney, D. A. (1989). On the role of the simplicity heuristic in language processing: Evidence from structural and inferential processing. *Journal of Psycholinguistic Research*, *18*(6), 553–562. https://doi.org/10.1007/BF01067159

Özer, D., & Göksun, T. (2020). Gesture Use and Processing: A Review on Individual Differences in Cognitive Resources. *Frontiers in Psychology*, *11*, 573555. https://doi.org/10.3389/fpsyg.2020.573555

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*(4), 688–704. https://doi.org/10.1006/jmla.2001.2826

Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R., & Kearney, J. K. (1989). Inhibiting The Lombard Effect. *Journal of the Acoustical Society of America*, *85*(2), 894–900. https://doi.org/10.1121/1.397561

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/s0140525x04000056

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*(6), 747–754. https://doi.org/10.1111/j.1467-7687.2007.00610.x

Pine, K. J., Gurney, D. J., & Fletcher, B. (2010). The Semantic specificity hypothesis: When gestures do not depend upon the presence of a listener. *Journal of Nonverbal Behavior*, *34*(3), 169–178. https://doi.org/10.1007/s10919-010-0089-7

R Development Core Team. (2011). *R: A language and environment for statistical computing.* https://www.r-project.org/

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*(4), 226. https://doi.org/10.1111/j.1467-9280.1996.tb00364.x

Rimé, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal variables. *European Journal of Social Psychology*, *12*(2), 113–129. https://doi.org/10.1002/ejsp.2420120201

Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*(4), 311–325. https://doi.org/10.1007/BF00991870

Riseborough, M. G. (1981). Physiographic Gestures As Decoding Facilitators: Three Experiments Exploring a Neglected Facet of Communication. *Journal of Nonverbal Behavior*, *5*(3), 172–183. https://0-link-springer-com.pugwash.lib.warwick.ac.uk/content/pdf/10.1007%2FBF00986134.pdf

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, *210*(5–6), 419–421. https://doi.org/10.1007/s00429-005-0039-z

Rossnagel, C. (2000). Cognitive load and perspective-taking: Applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, *30*(3), 429–445. https://doi.org/10.1002/(sici)1099-0992(200005/06)30:3<429::aid-ejsp3>3.0.co;2-v

Schober, M. F., & Clark, H. H. (1989). Understanding by Addressees and Overhearers. *Cognitive Psychology*, *21*(21). http://0-ac.els-cdn.com.pugwash.lib.warwick.ac.uk/001002858990008X/1-s2.0-001002858990008X-main.pdf?_tid=1c97e378-7e97-11e7-81fe-00000aacb361&acdnat=1502457494_24e9a31de241a7b9f368d753d015d208

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of Retrieval as Information: Another Look at the Availability Heuristic. *Journal of*

Personality and Social Psychology, *61*(2), 195–202. https://doi.org/10.1037/0022-3514.61.2.195

Shintel, H., & Keysar, B. (2009). Less Is More: A Minimalist Account of Joint Action in Communication. *Topics in Cognitive Science*, *1*(2), 260–273. https://doi.org/10.1111/j.1756-8765.2009.01018.x

Slonimska, A., Ozyurek, A., & Campisi, E. (2015). Ostensive signals: markers of communicative relevance of gesture during multimodal demonstrations to adults and children. *4th GESPIN-Gesture & Speech in Interaction*.

Smithson, L., & Nicoladis, E. (2013). Verbal memory resources predict iconic gesture use among monolinguals and bilinguals. *Bilingualism*, *16*(4), 934–944. https://doi.org/10.1017/S1366728913000175

Snow, C. E. (1972). Mothers' Speech to Children Learning Language. *Child Development*, *43*(2), 549. https://doi.org/10.2307/1127555

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*, *33*(1), 115–125. https://doi.org/10.1111/j.1551-6709.2008.01006.x

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, *41*(1), 31–57. https://doi.org/10.1080/08351810701691123

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

Streeck, J. (2008). Depicting by gesture. *Gesture*, *8*(3), 285–301. https://doi.org/10.1075/gest.8.3.02str

Streeck, J. (2009). Forward-gesturing. *Discourse Processes*, *46*(2–3), 161–179. https://doi.org/10.1080/01638530902728793

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Suppes, A., Tzeng, C. Y., & Galguera, L. (2015). Using and Seeing Co-speech Gesture in a

Spatial Task. *Journal of Nonverbal Behavior*, *39*(3), 241–257. https://doi.org/10.1007/s10919-015-0207-7

Theron-Grimaldi, S. (2021). *A handy effect: the communicative benefits of gestures on memory through an investigation of the encoring-recall modality congruency*. University of Warwick.

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

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. In *Contemporary Educational Psychology* (Vol. 28, Issue 2, pp. 187–204). Academic Press. https://doi.org/10.1016/S0361-476X(02)00007-3

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology*, *114*(4), 591–600. https://doi.org/10.2307/1423612

Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage*, *47*(4), 1992–2004. https://doi.org/10.1016/j.neuroimage.2009.05.066

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Winner, T., Selen, L., Murillo Oosterwijk, A., Verhagen, L., Medendorp, W. P., van Rooij, I., & Toni, I. (2019). Recipient Design in Communicative Pointing. *Cognitive Science*, *43*(5), e12733. https://doi.org/10.1111/cogs.12733

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

Yngve, V. H. (1970). On getting a word in edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, 567–578. https://doi.org/10009705656

Appendix 1 – Gesture Coding Manual

This coding manual is based on the one used in Chu et al. (2014). The large majority of it is the same as Chu et al. (2014), but the representational, beat, and representational/listener-indexing sub-sections of the gesture classification section have been updated.

GESTURE CODING MANUAL

INDIVIDUAL DIFFERENCES PROJECT

B. Hannah, M. Chu & S. Kita

Additional notes added by Lucy Foulkes 24/08/11

Edited by Jacob Barker 01/06/17 & 11/02/2019

GESTURE CLASSIFICATION

The following gesture classification is based on McNeill (1992). What is called "depictive gestures" here include iconic and metaphoric gestures in McNeill (1992).

1. REPRESENTATIONAL (REP)

These are depictive and deictic gestures. In depictive gestures, the form of gesture can be interpreted, in the context of concurrent speech, as depicting a physical action by a character (human) or perceptual properties or movements of an object (e.g. beans, the whole human body) or an entity (e.g. secret). Depictive gestures can also metaphorically depict abstract concepts (see below). The hand movement in depictive gestures resemble the referent concept. In deictic gestures, gestures indicate a location or establish meaning in a location in space.

There are several types of rep gesture (although this list is not exhaustive):

a) Hand-as-a-concept

A hand represents a concept (e.g. two hands alternately move up and down, to represent two friends in conflict. Each hand represents a friend.)

b) Concept-on-a-palm

The speaker may position their hand with the palm (clearly) upwards, as if s/he is holding a concept or an abstract entity on the palm (e.g. "He's got an idea").
The speaker is not explicitly conveying the idea to the listener (this would be a rep/listener-indexing gesture). The holding or possession can be metaphorical or implied (e.g. "she is very happy" => she has the emotion; "she knows what's happening" => she has knowledge). In the clearest cases, the concurrent speech refers to actual having or holding. This is typically unimanual.

c) Shape-of-a-concept/Holding-a-concept

The speaker may position their hands in one of two ways:

i) as if to indicate the shape of an object (e.g. two hands open up to form a spherical shape), which establishes the presence of a concept.

ii) as if to indicate holding of an object (e.g. two hands are oriented diagonally upward and also facing each other), which not only establishes the presence of a concept but also suggests possession of, and control over, the concept.

d) Concept-changing-shape

The speaker may move their hands as if the concept they are holding changes its shape (e.g. two hands look as if they are holding a concept, and the distance between them becomes bigger as if the concept is expanding).

e) Concept-manipulation

The speaker may move their hands as if to indicate s/he is moving a concept to a particular location, or manipulating the shape of a concept (e.g. two hands repeatedly push a concept down, as if kneading).

N.B. In the Individual Differences project, we do not distinguish between different types of representational gestures.

2. REPRESENTATIONAL UNCLEAR (REP UNCLEAR)

These gestures look like they are representational, but have no clear representation in the concurrent speech. They will often, but not always, look the same as the previous or following (representational) gesture.

3. BEAT

This gesture is a small biphasic movement (up and down), typically involving a wrist movement. There may be several up-and-down movements in succession. The hand shape is open and lax (none of the fingers are singled out). The gesture happens in the periphery of the gesture space (e.g. on the lap).

In some cases, the gesture only has some features of a beat gesture (e.g. the gesture has a loose hand shape and is in the periphery, but there is no biphasic movement). In this situation, you should still code it as beat – unless you can find **meaning** in the **form** of the hand shape. For example, if the hand moves slightly forward and the subject is talking about ‘progressing’, then this should be coded as rep, not beat.

N.B. It is not sufficient to see meaning in a beat gesture (and call it rep) simply when a word coincides with the stroke movement. Otherwise, almost all beat gestures would be coded as rep gestures. There must be more evidence that the form relates to meaning, e.g. placement to the side, moving forwards, the hand shape. If an interpretable meaning can be established, then the gesture should be coded as rep.

N.B. When determining if a gesture contains meaning, surrounding gestures can provide meaning. For instance the gesture can reference locations that previous locations have established meaning for.

N.B. a gesture where the participant claps their hands on their legs should be counted as a beat gesture.

4. PALM-REVEALING EMBLEM (PALM-UP EMBLEM)

For this gesture, the hand orientation is typically palm up, or the hand turns to reveal the palm. The hand could also partially turn to a position that reveals the palm, but does not complete this rotation (e.g. palm down to palm sideways). It is often accompanied by a shoulder shrug and/or facial expressions (eyebrow raise). It can be unimanual or bimanual symmetrical.

The general motivation for the gestural form is "empty-handedness": the speaker has nothing to show or share. There are several scenarios in which this gesture may be used:

a) The speaker is **uncertain** about what they are talking about- e.g. a concept, decision, or definition. The gesture accompanies speech such as, “it’s like”, “kind of”, “I don’t know”, or the speaker may stutter, make filled pauses (e.g. “um”), or repeat words.
b) The speaker has a lack of knowledge, possibly with overtone of lack of interest, such as, "whatever", "I guess", “It doesn’t matter”.
c) At the end of a description or a journey. It typically accompanies meta-level speech such as, "that's all", "that's all it did", “there’s nothing more to say”, or may be produced with no speech at all.
d) When the speaker considers something to be obvious, e.g. “She has to talk to Peter”, “Of course it’s a bad idea.”
e) The speaker is expressing resignation, as though to say, “I wouldn’t have a choice”, “That’s just how it is”.
f) In a turn taking situation, when the speaker has finished, suggesting, “I am done.” (N.B. This doesn’t occur in this project.)

Notes

- The palm-revealing gesture can be produced in any of the above scenarios without speech. In this case, the speech before and/or after the gesture will generally demonstrate the meaning of the palm-revealing gesture.
- If a shoulder shrug occurs as a gesture by itself, this should always be coded as palm-revealing, regardless of the speech. This be coded as a non-hand body part gesture (torso).

5. LISTENER-INDEXING

These are "interactive gestures" (Bavelas et al., 1992) which point to the listener. The handshape can be an index-finger point, but more often, it is a loose handshape. The fingers do not necessarily point to the listener. The handshape should move towards the listener.

Examples of listener-indexing gestures:

a) The speaker points at the listener, as s/he says "you are absolutely right!"/"As you said earlier..."

b) The speaker moves the hand towards the listener in a loose handshape, as she says "Don't you think?".

N.B. These other gestures are very rare on this project

6.REPRESENTATIONAL/LISTENER INDEXING (REP/LISTENER)

This type of listener indexing gesture is also known as a conduit gesture. It can be interpreted as presenting an idea or message on the palm to the listener. The speaker must have **eye contact** (see N.B. below) with the listener at the **end** of the gesture. The speaker's hand must move predominantly **towards** the listener (not predominantly vertically or sideways), and their **palm** must be revealed to the listener (or their hands rotate partially so that if the rotation continued the movement their palm would be revealed). The gesture should also be located in the centre of gesture space between the speaker and listener, not off to the left or right side of space. An indicator (though not a requirement) is the orientation of the hand. The hand should be oriented so the fingers are pointing in the direction of the listener.

The speaker often intonates their speech as if to emphasize the concept or idea that they want to express to the listener. The speaker may also show other nonverbal cues for the orientation towards the listener such as leaning forward (towards the listener) and nodding. Unlike palm-revealing gestures, a listener-indexing gesture is produced when the speaker presents a clearly formulated idea without any uncertainty.

N.B. A clear distinction between concept-on-a-palm representational gestures should be made. Conduit gestures should display additional movement beyond what is necessary for a concept-on-a-palm gesture. Even if the wrist rotates so that the hand moves towards the listener, evidence of extra effort is required to state that the gesture is interacting with the listener. This could be an additional rotation of the hand to face the palm towards the listener beyond what is required for a concept-on-a-palm gesture. The speaker could also change their handshape so that the fingers are angled downward and facing the listener. If the fingers are pointing at the listener this means the finger tips should be lower than the palm. If the fingers are pointing to the side, then the finger closest to the listener should be lower than the next closest finger and so on.

N.B. When a gesture meets the criteria for a conduit gesture, but has an alternative representational meaning other than a concept being passed from the speaker's palm to the listener it should be coded as representational only.

N.B. If a representational gesture has a conduit gesture incorporated into it, when it should be classed as a conduit gesture. (Ie performing a characters action but changing the action of one hand at the end to be a conduit gesture.

N.B. When there is a visibility manipulation present during the experiment, then eye contact removed from the criteria for all conditions, not just those where eye contact is directly blocked.

7.OTHER NON-REPRESENTATIONAL (OTHER NON-REP)

Describe what the gesture is in the comment tier:

a) <u>Emblem</u>

Examples include the "OK-sign" and "thumbs-up". Another commonly seen one is the "hedging emblem", in which the hand orientation is typically palm-down (palm orientation may vary), and the hand wavers. The gesture can be interpreted as hedging, and the speech also refers to uncertainty, for example, "maybe".

N.B. When participants clasp their hands together with an audible clap or actually clap their hands together, the gesture should be classified as an other non-rep gesture.

b) <u>Metacognitive</u>

This is a repetitive movement that indicates that the speaker is thinking. It may indicate frustration. The gesture in this category must resemble continuous multiple finger tapping or continuous multiple rotary movement of a hand ("what was the word for it?").

Notes

- Sometimes, a gesture will be a mixture of rep and other non-rep. In this case, code the gesture as rep. As long as you can see some meaningful representation in the gesture, always code this as rep.
- When the speaker acts out the gestures of a particular character in the story, the gesture should be coded at face value and then put a code (XCH) in the gesture comment tier.

8.UNCLEAR

These include the following:

a) Abandoned gestures: gestures that were prematurely stopped before completion. They normally only consist of a preparation and a retraction phase, without any stroke.
b) Gestures that cannot be coded as any of the category above.

GESTURE SEGMENTATION

1. Most gestures consist of three phases: preparation, stroke and retraction. We are interested in the **stroke**, and this is the bit which should be segmented. Note: not all gestures consist of these! E.g. a gesture may have no preparation phase if it leads on directly from the previous gesture.

2. The stroke is typically the fastest and clearest part of the gesture. It generally also coincides meaningfully with the speech.

3. Rachel and Valentina - Sometimes it is difficult to decide whether or not to code a hand movement as a gesture. In this case, it is better to code it than to leave it and potentially miss a gesture. It is better to code too many than

too few, as someone will check your codings. Make a note in the "gesture comment" column for the codings you are unsure about.

4. Sometimes it is difficult to decide how to divide up hand movements into gestures. Generally, if there is a pause between movements, or a change of direction or speed, this indicates the start of a new gesture. It is also helpful to listen to the speech, as a new word or topic might indicate that it is a new gesture.

5. Exactly where you segment the gesture is not crucial, so do not worry too much about identifying exactly where the gesture starts or finishes. This is important for some projects (e.g. looking at speech/gesture synchronicity), but we are only looking at number of gestures, so don't spend too much time on this if it's difficult.

SALIENCE CODING

Gesture salience is categorized into the following four categories:

a) Finger(s) only: When only the finger(s) move, but the hand remains still
b) Hand: When the hand(s) moves at the wrist in an up/down or left/right plane, but the forearm remains still. Or, the hand rotates and the forearm rotates in consequence
c) Forearm: When the forearm(s) move in an up/down or left/right plane, but the upper arm remains still. If the forearm rotates, as a result of the hand rotating, code this as hand only.
d) Whole arm: When the whole arm, including the upper arm moves in an up/down or left/right plane

It is crucial to consider that, whenever one part of the arm moves, a bigger, connecting arm part may move slightly as a natural consequence of this. Always take this into account when coding salience. If the bigger part moves only slightly, and not in an up/down or left/right plane, ignore this movement and code the smaller arm part. BUT, if the bigger part is moving in an up/down or left/right plane, or the movement is more than a natural consequence of the smaller part moving, code the bigger part. Sometimes, it is difficult to make this decision, so a careful judgement must be made.

HEIGHT CODING

Gesture height is categorized into the following three categories:

a) lower than waist
b) between waist and chin
c) above chin

To code height, you first need to identify where the participant's waistline is. The waist line lies between the hips (i.e., the top of a person's trousers) and the bottom of the ribcage.

Make a careful judgement about this, and then mark the waistline using a post-it note on the computer screen.

To make a judgement about whether a person's gesture is above or below the chin line, see where the chin is in that particular gesture. If it helps, hold up a post-it note to see a straight line from the bottom of the subject's chin.

Height coding must be based on the highest point of the hand/finger in the gesture. If this falls on or below the waistline that you have marked, code "lower than waist". If it is above this, but on or below the chin line, code "between waist and chin". If the highest point of the hand reaches above the chin line, code "above chin".

HANDEDNESS CODING

In this tier, we code which hand is involved in the gesture. There are three choices: LM (left movement), RM (right movement) and LM-RM (left and right movement, i.e. both hands are involved in the gesture). Ignore the "hold" categories (e.g. LH); these are not relevant for this project.

To code handedness, decide which hands are *meaningfully* involved in the gesture. Sometimes, both hands are moving, but only one is moving meaningfully. In this case, code only RM or LM. If both are contributing meaningfully to the gesture, then code LM-RM. A good way to decide is to ask, "Which hand(s) is representing something?"

References

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*(2), 694–709. https://doi.org/10.1037/a0033861

Appendix 2 – Head Movement Coding Supplement

GESTURE CODING MANUAL

Head movement Supplement

J. Barker & S. Kita

HEAD MOVEMENT CLASSIFICATION

1. *REPRESENTATIONAL (REP)*

Depictive head movements which can be interpreted, in the context of concurrent speech, as depicting a physical action by a character (human) or perceptual properties, or movements of an object or an entity. Deictic head movements can indicate a location or establish meaning in a location in space.

Representational head movements can co-occur with representational hand gestures. Head movements and hand gestures can convey related information.

a) Representational head movements will typically refer to a location in space or a movement from one place to another.
b) Nodding to highlight certain words should not be coded as representational head movements unless there are other arguments to support the movement being representational (e.g. a larger head movement when saying “she hit the cat with her umbrella”).
c) Speakers do not always fixate on the addressee, for example when searching for a word or between sentences. Speakers often alternate between averting their gaze and fixating back on the addressee. Head movements to avert their gaze away from the addressee, or movements to fixate back on the addressee, should not be coded as representational head movements unless there are other arguments to support the movement being representational (e.g. moving the head quickly to the side when saying “the cat came flying out of the window”).
d) Information from the stimuli can help clarify if a head movement is representational. Gestures often reflects the lateral (left vs right) direction of the movement in the stimuli (English speakers 77% congruent direction, 23% non-directional, 0% incongruent direction (Kita & Ozyurek, 2003). If head movements reflect the movement in the cartoon, then the head movement is likely to be representational.
e) Surrounding speech and gestures can be used to help interpret head movements. Head movements that move in the same direction as co-occurring gestures are likely to be representing the same information (and therefore representational).
f) Surrounding head movements can provide additional information. Speakers often produce nods for emphasis when talking. If a nod occurs when a participant is referring to something moving downwards in speech, only a larger movement (distinct from previous nods) should be coded as representational.

2. INTERACTIVE (LISTENER-INDEXING)

These are head movements which relate to the listener. These include head nods (e.g., in the context of agreeing or acknowledgement) or head shakes (e.g., in the context of disagreeing, disbelief, and intensification, uncertainty, McClave, 2000, McClave et al., 2007) or head tilts (e.g., in the context of the speaker's uncertainty, similar to palm revealing gestures).

3. NODS FOR EMPHASIS (BEAT)

These are head movements that speakers produce to emphasise a word. When speakers nod for emphasis, speakers typically move their head down and then up again. This movement will typically co-occur with a word the speaker is emphasising. There can be several nods for emphasis in succession.

HEAD MOVEMENT SEGMENTATION

Head movements should be segmented in the same manner as gestures. The “stroke” of the head movement is typically the fastest and clearest part of the gesture. It generally also coincides meaningfully with the speech.

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.

McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics, 32*(7), 855-878. doi:https://doi.org/10.1016/S0378-2166(99)00079-X

McClave, E., Kim, H., Tamer, R., & Mileff, M. (2007). Head movements in the context of speech in Arabic, Bulgarian, Korean, and African-American Vernacular English. *Gesture, 7*(3), 343-390. doi:https://doi.org/10.1075/gest.7.3.04mcc

Appendix 3 – Social Dilemmas

Stimuli used in the social dilemma tasks

Stories 1-4 are taken from Chu et al. (2014) and are unchanged. Stories 5 and 6 are new stories written for this study.

Story 1

Hannah and **Louise** have been best friends for many years and have just started university together. **Hannah** is enjoying the freedom of university and particularly enjoys socialising. **Louise** is missing home and is not enjoying her university experience; she hasn't made any new friends and is very dependent on **Hannah**. **Hannah** feels guilty for having fun with her new friends knowing **Louise** is miserable. **Hannah** tries to encourage **Louise** to go to parties with her, but knows her new friends think **Louise** is geeky and do not like her. In recent weeks, **Hannah** has started to feel torn as both her new friends and **Louise** have been inviting her to go out increasingly often. She enjoys the company of her new friends but feels obliged to spend time with **Louise**.

- If you were in **Hannah**'s situation, what would you do and why?
- What do you think **Louise** would feel about your decision?
- What do you think your new friends would feel about your decision?

Story 2

Amy and **Peter** have been going out for two years. They are both best friends with **Katie**. **Amy** always tells **Katie** how lucky she is to have **Peter** but recently **Katie** has noticed **Peter** getting very close to another girl. **Peter** assures **Katie** that he and the other girl are just friends and that he loves **Amy** very much. **Katie** has since seen **Peter** and the other girl acting as more than just friends on more than one occasion but is not totally sure what to make of the situation. **Katie** is not sure whether to tell **Amy** or not.

- If you were in **Katie's** situation, what would you do and why?
- What do you think **Peter** would feel about your decision?

- What do you think **Amy** would feel about your decision?

Story 3

Anne has been at University for just under a year. She has struggled to make many other friends beside her best friend, **Belle**. **Belle** is very popular and knows a lot of people. **Belle** is having a huge 18th birthday party, which everyone will be at. **Belle** tells **Anne** that this is the perfect opportunity for her to meet new people and have fun. However **Belle's** 18th birthday party is on the same day as **Anne's** favourite **aunt's** 40th birthday party. The **aunt** has personally called **Anne** to invite her to the party.

- If you were in **Anne's** situation, what would you do and why?
- What do you think **Belle** would feel about your decision?
- What do you think your **aunt** would feel about your decision?

Story 4

Mary is an undergraduate student in a university and her department is holding a ball. Her best friend **Tom**, whom she has known since the start of university, really wants to go to the ball, and has hinted at wanting to go but having no one else to go with. **Mary** has recently met **Tom's** friend **Peter**, who she really likes. **Mary** really wants to go on a date with **Peter** and figures the ball is a good excuse to ask him out. Though **Peter** has stated he doesn't really enjoy these sorts of social events, this may be the last chance as he is going abroad soon as an exchange student. Because **Tom** and **Peter** are good friends with each other, if **Mary** asks one out, the other will find out. Tickets are sold in pairs, and they need to be bought in the next few hours, so **Mary** needs to ask one of them soon.

- If you were in **Mary's** situation, what would you do and why?
- What do you think **Tom** would feel about your decision?
- What do you think **Peter** would think about your decision?

Story 5

Jess is getting married soon, and needs to pick a maid of honour to help her with all the arrangements. **Hazel** and **Lucy** are her two closest friends and would make suitable options. **Hazel** is her oldest friend, and knows how **Jess** want's her wedding to be, and would always ask **Jess**' preference. However she can be very lazy and is not very organised. **Lucy**, in contrast, is very motivated to help with the arrangements and is very organised. She has arranged weddings before and those had been great successes. However **Lucy** likes to do things in her own way, and may not always listen to **Jess**' wishes. Both **Hazel** and **Lucy** had expressed interest in being **Jess**' maid of honour, and she needs to make a decision.

- If you were **Jess**, what would you do and Why?
- What do you think **Hazel** would feel about your decision?
- What do you think **Lucy** would feel about your decision?

Story 6

Susan is graduating from her undergraduate degree, and has two tickets to her graduation ceremony. Her parents are separated and not on speaking terms, and would not consider going together. Both would bring her younger brother to the ceremony with the other ticket. **Susan** is on good terms with both of her parents, but was brought up by her **Father** and would consider herself closer to him as a result. However her **Mother** has helped her financially throughout the degree, both for the tuition and for some of her living expenses. **Susan** is immensely grateful for her **Mother's** contribution as it had taken the burden of worrying about finances out of her time at university.

- If you were **Susan**, who would you take and Why?
- What do you think Susan's **Father** would feel about your decision?
- What do you think Susan's **Mother** would feel about your decision?

References

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*(2), 694–709.

https://doi.org/10.1037/a0033861

Appendix 4 – Example foreground, background and absent gesture frames

Images to help demonstrate the stimuli used in Chapter 5

Figure 1

An example of a visually prominent gesture. The speaker waves, moving their forearm in wide motions left and right in a prominent location in gesturer space.

Figure 2

An example of a non-visually prominent gesture. The speaker waves, moving their hand left and right in a peripheral location in gesturer space.

Figure 3

An example of an absent “gesture”. The speaker simply does not produce a gesture in this condition. There is no demonstrative phrase, nor gaze fixation.

Figure 4

An example of a gaze-fixated gesture. The speaker waves, moving their hand in left and right in a prominent location in gesturer space. The speaker begins fixating on the gesture at the beginning of the stroke phase and ends fixating on the gesture at the end of the stroke phrase.

Figure 5

An example of a non-gaze-fixated gesture. The speaker waves, moving their hand in left and right in a prominent location in gesturer space. The speaker does not fixate on the gesture.

Appendix 5 – Drawing Target Event Coding

Coding scheme used for marking participant's drawings as containing the target information or not. Created by (Theron-Grimaldi, 2021).

What we want to achieve is to code if participants mentioned the target information correctly when they remembered the event. Using this coding scheme, we want to ensure that participants did understand what event they were asked to remember and therefore had the opportunity to mention the target information.

To be counted as being a remembered event, the response must first contain the correct protagonist(s). In addition to that, the participant needs to at least mention the correct action or an object relevant to the scene and the protagonist(s)'s action. To give an example, in the event « the cat is rolling through the air », the participant needs to mention the cat and the fact that he is moving in the air. Some specifications need to be mentioned. First, the action or feature mentioned do not need to be exact, but at least similar (eg. the cat 'going up', instead of 'climbing'). Moreover, if participants clearly muddled up to another event, the event will not be considered as being remembered, even if it met the criteria mentioned above.

When the event is considered being remembered, the second part of the coding is focusing on the target information. The target information (eg: the location where the cat crashed) is either specified or underspecified (eg: « The cat crashed. »). When specified, we analyse if it is mentioned correctly (eg: « the cat crashing under the bird's window ») or in a contradictory manner (eg: « The cat crashing next to the bird's window. »). When a target information is not mentioned accurately enough (eg: « The cat crashing near the bird's window. »), it is neither considered correct nor contradictory.

Table 1

A list of the target events, target information, and drawing criteria for all three experiments of Chapter 5

Cartoon	Event ID	Event	Target information	Target Information Drawing Correct
1. Binoculars	1.1	The bird waving	waving	A clear hand or wing raised up, with or without an arrow

1. Binoculars	1.2	The cat jumping over a car	jumping over	Any arrow or the cat above the car
1. Binoculars	1.3	The cat rolling through the air	rolling	An arrow that turns
1. Binoculars	1.4	The cat removing a banana skin on his shoulder	on his shoulder	A banana on the cat's shoulder
2. Drainpipe	2.1	Cat climbing inside the drainpipe	inside	The cat and/or arrows inside the drainpipe
2. Drainpipe	2.2	The bird escaping from the top of the cage	from the top	Arrow going from the incurved top part of the cage
2. Drainpipe	2.3	The grandma slapping the cat	slapping	One hand or the umbrella touching or being next to the cat
2. Drainpipe	2.4	The grandma waving the umbrella in the air	in the air	The grandma waving the umbrella to the air
3. Tramline	3.1	The cat walking on a line balancing	balancing	An arrow or object showing he is balancing
3. Tramline	3.2	The grandma tugging on the bell	tugging	An arrow going up and/or down
3. Tramline	3.3	The cat spirals down the pole	spirals	An arrow with a spiral
4. Monkey	4.1	The cat reappearing through the bushes	through the bushes	The cat reappearing inside the bushes
4. Monkey	4.2	The cat sneaking through the top of the window	the top	Window being opened at the top
4. Monkey	4.3	The cat scooping up the bird	scooping up	The hat facing upward with the bird inside it
4. Monkey	4.4	The cat scratching his body	scratching his body	The cat having at least one arm bent towards its body
4. Monkey	4.5	The cat holding the cup with both hands	with both hands	The cat using two arms to hold the cup
4. Monkey	4.6	The cat having a bump on the side of the head	on the side	A bump on the side of the head
5. Bowling	5.1	The bird throwing the ball to the side	to the side	An arrow or both hands going sideway, with the drainpipe next to the bird or window (it is needed to check if the drawing is on the profile or front view)
5. Bowling	5.2	The cat thrown out of the pipe bouncing	bouncing	One or multiple arrows going in an up/down motion
5. Bowling	5.3	The cat rolling down the street zigzagging	zigzagging	Arrows going left and right
6. Bellboy	6.1	The cat reaching some letters above him	above him	The cat with a hand or arrow going up
6. Bellboy	6.2	The cat knocking on the door	knocking	A hand touching the door with a "knock" motion

6. Bellboy	6.3	The cat grabbing the cage with both hands	with both hands	Both hands touching the birdcage; the birdcage close to the cat upper body
6. Bellboy	6.4	The grandma jabbing the cat with the umbrella	jabbing	An arrow going fowards; the umbrella pointing towards the car
7. Swing	7.1	The bird swinging on her cage	swinging	An arrow going to the side
7. Swing	7.2	The cat crashing under the bird's window	under	A window and/or the bird placed above the cat
7. Swing	7.3	The cat smashing on his head	on his head	The head facing the ground
8. Catapult	8.1	The cat holding the square weight	square	The weight with angular forms
8. Catapult	8.2	The cat throwing the weight over his head	over his head	The weight or arrows being above the cat
8. Catapult	8.3	The cat grabbing the bird with two fingers	with two fingers	The cat holding the bird with two fingers only

Table 2

Descriptions of the visually prominent and non-visually prominent gestures in Experiment 1 of Chapter 5

Event ID	Target information	Visually prominent Gestures	Non-Visually prominent Gestures
1.1	waving	Forearm and hand moved in wide motions left and right in a prominent location.	Hand moved in small left and right movements on lap
1.2	jumping over	Whole arm with pointed finger traced arc from one side of the body to the other (going up then down)	Hand with pointed finger traced arc from one side of a leg to the other (going up then down)
1.3	rolling	Whole arm with pointed finger traced a spiral from one side of the body to the other	Hand with pointed finger traced arc from one side of a leg to the other (going up then down)
1.4	on his shoulder	A whole arm sweeping motion starting with a hand near the neck and moving along the length of the shoulder and outwards	A hand sweeping motion starting with a hand in the middle of the shoulder and moving along the length of the shoulder and stopping
2.1	inside	One hand forms the outside of a cylinder at upper chest height, the other moves up inside and through the semi-circle while pointing upward. The movement goes from the stomach area to the head area.	One hand forms the outside of a cylinder at on the speaker’s lap. The other moves up inside the semi-circle while pointing upward. The movement goes from the lap area to the stomach area.

2.2	from the top	One hand forms the outside of a semicircle at upper chest height, the other moves up inside, through, and round the semi-circle while pointing in the direction of motion. The movement goes from the stomach area to the head area.	One hand forms the outside of a semicircle at lap height, the other moves up inside, through, and round the semi-circle while pointing in the direction of motion. The movement goes from the lap area to the stomach area.
2.3	slapping	The speaker makes a sweeping motion with their whole arm. The back of the hand facing the direction of travel. The movement starts on one side of the body and ends far on the other side. The movement is at chest height.	The speaker makes a sweeping motion with their hand. The back of the hand facing the direction of travel. Only the wrist rotates so the movement does not go far. The movement is at lap height.
2.4	in the air	The speaker closes his fist as if holding an umbrella. The speaker makes back and forth motions over their head going from one side of their body to the other.	The speaker closes his fist as if holding an umbrella. The speaker makes back and forth motions to one side of their body at lower chest height. The total distance is only a couple of hand widths.
3.1	balancing	The speaker outstretches their (whole) arms to either side and alternates raising one hand while lowering the other. The hands go from head to stomach height	The speaker moves their forearms to either side and alternates raising one hand while lowering the other. The hands go from head to lap to stomach height
3.2	tugging	The speaker closes their fist (as if around a cord) at head height and pulls downward with their whole arm several times to lower chest height	The speaker closes their fist (as if around a cord) at stomach height and pulls downward with their forearm several times to lap chest height
3.3	spirals	The speaker moves their arm and hand downwards rotating their forearm/hand to create a spiral path downwards. The movement goes from head height to stomach height.	The speaker moves their finger downwards rotating their hand to create a spiral path downwards. The movement goes from stomach height to lap height.
4.1	through the bushes	The speaker touches the backs of their hands (technically fingers) together at head height and moves each hand outwards with a whole arm movement. The hands end up on either side of their body.	The speaker touches the backs of their hands (technically fingers) together at stomach height and moves each hand outwards with a forearm movement. The hands end up two hand widths apart.
4.2	the top	One hand forms a flat vertical plane, fingers upward, with the heel of the hand (little finger) facing the camera. The hand is held at upper chest height. The other hand moves in an arc over the first hand reaching head height before moving down on the other side.	One hand forms a flat vertical plane, fingers upward, with the heel of the hand (little finger) facing the camera. The hand is held in the speakers leg at lap height. The other hand moves in an arc over the first hand reaching stomach height before moving down on the other side.

4.3	scooping up	The speaker forms his hand as if on the outside of a cylindrical object and scoops with the hand from upper chest height on one side of the body, down to stomach height, and back up to chest height on the other side of the body with a whole arm movement.	The speaker forms his hand as if on the outside of a cylindrical object and scoops with the hand from stomach height.
4.4	scratching his body	The speaker bent their arms so the hands were touching their sides at stomach height. The hands are moved up and down still touching the speaker's sides up to the armpits and back down several times. The whole arm is used to move the hands up and down	The speaker bent their arms so the hands were touching their sides at lap height. The hands are moved up and down still touching the speaker's sides up to stomach height and back down several times. Wrist rotations is used to move the hands up and down
4.5	with both hands	The speaker moves his hands together with a whole arm movement at upper chest height to form a large circle as if holding a cylinder.	The speaker moves his hands together at lap height to form a circle as if holding a cylinder.
4.6	on the side	The speaker puts their hand to the side of their head with the fingers splayed out to make a cone shape. The speaker moves his hand away from their vertically while closing the fingers together to a point 3 – 4 hand widths away from the side of the head.	The speaker taps the side of their head with two fingers.
5.1	to the side	Two hands face each other as if holding a bowling ball between them at upper chest level. The speaker moves both arms (whole arm movement) to the same side, pushing with one hand before moving the hands apart while moving sideways (to let go of the ball)	Two hands face each other as if holding a bowling ball between them on the speaker's lap. The speaker moves both hands (wrist movement) to the same side, pushing with one hand before moving the hands apart while moving sideways (to let go of the ball)
5.2	bouncing	While pointing, the speaker traces a series of arcs from one side of their body to the other side of their body at upper chest level. The movement involves the whole arm	While pointing, the speaker traces a series of arcs from one side of a leg to the other side of their leg, while resting on their lap.
5.3	zigzagging	While pointing, the speaker traces a series of zigzags from head height to stomach height. The movement involves the whole arm	While pointing, the speaker traces a series of zigzags on their lap downwards from stomach to lap height. The movement involves only the wrist/hand
6.1	above him	One hand forms a horizontal line at shoulder height while the other reaches over it from the front and	One hand forms a horizontal line at stomach height while the other reaches over it from the front and makes a grabbing motion (the forearm moves).

		makes a grabbing motion (the whole arm moves).	
6.2	knocking	The speaker makes a fist and moves the hand forwards at head hight repeatedly with a whole arm movement.	The speaker makes a fist and moves the hand forwards at lap hight repeatedly with a wrist movement.
6.3	with both hands	The speaker wraps their arms around a space in front of himself at upper chest level. It is a whole arm movement.	The speaker moves their hands, so the hands are overlapping with the backs of the hands facing the camera at stomach height. It is a forearm movement.
6.4	jabbing	The speaker makes a fist as if holding the handle of the umbrella. The speaker moves their whole arm so the fist is propelled forward repeatedly at chest height.	The speaker makes a fist as if holding the handle of the umbrella. The speaker moves their forearm so the fist is propelled forward repeatedly at stomach height.
7.1	swinging	The speaker uses their finger to motion a downward moving arc from one side of their body to the other using a whole arm movement at chest level (down to stomach and back up to chest).	The speaker uses their finger to motion a downward moving arc from one side of to another using a wrist movement on their lap.
7.2	under	One hand forms a vertical plane at chest height, the other moves a pointing finger towards and then under the other hand. The whole arm moves.	One hand forms a vertical plane at stomach height, the other moves a pointing finger towards and then under the other hand. The forearm moves.
7.3	on his head	The speaker moves their hand up toward the side of their head and moves sharply towards the temple with the flat of the hand	The speaker moves their hand slightly near the side of their head without a hand movement or sharp motion at the end.
8.1	square	Using two hands, the speaker motions out the outline of a square with pointing fingers. The gesture is at upper chest level making a large square shape with whole arm movements.	Using two hands, the speaker motions out the outline of a square with pointing fingers. The gesture is at stomach level making a small square shape with wrist movements.
8.2	over his head	The speaker lifts up his arms and both hands behind their head, on either side. Then the speaker swiftly brings the hands up and over their head (similar to a football throw-in) before moving the hands apart	The speaker lifts up their arms and both hands next to their head, on either side. The speaker flicks their wrists to move the hands upwards and apart.
8.3	with two fingers	The speaker moves their hand toward the camera with a whole arm movement and pinches with two fingers at head height.	The speaker pinches two fingers together with their hand on their lap.

Figure 1

An example of a participant's drawing (Event ID 1.1) meeting the target information criteria (A clear hand or wing raised up, with or without).

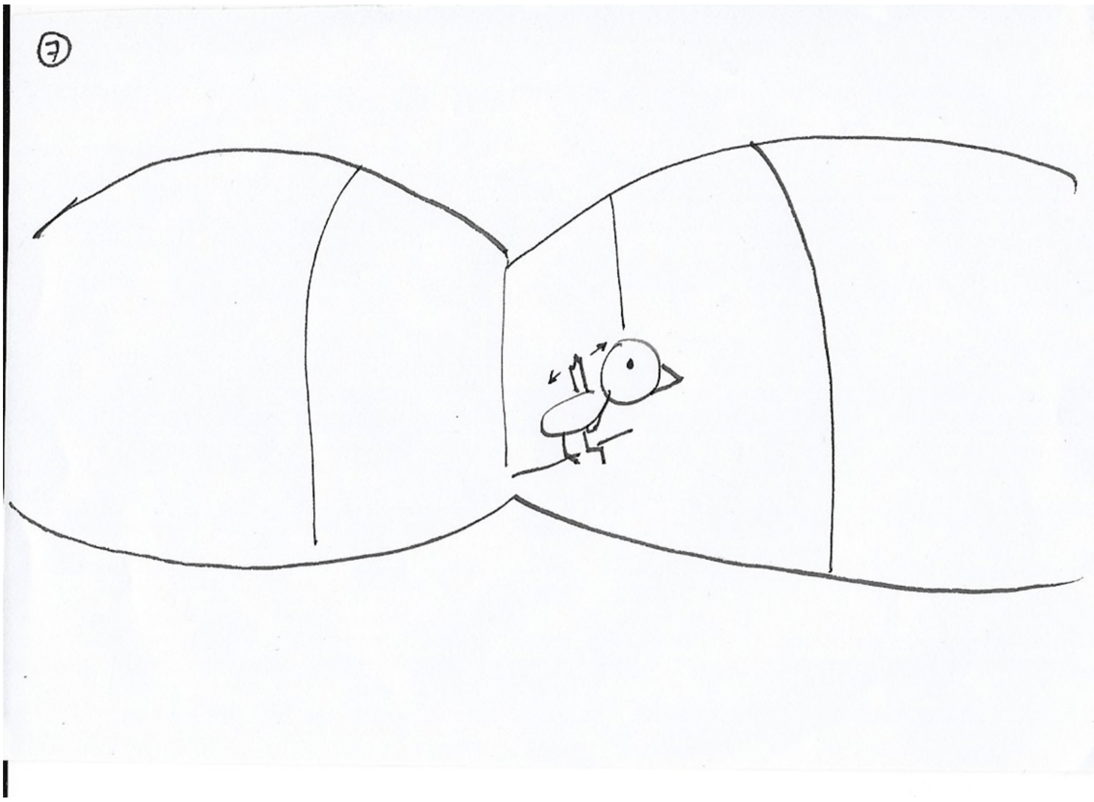


Figure 2

An example of a participant's drawing (Event ID 1.1) failing to meet the target information criteria (A clear hand or wing raised up, with or without).