

**What co-speech gestures do:  
investigating the communicative role of  
visual behaviour accompanying  
language use during reference in  
interaction**

Jack J. Wilson

Submitted in accordance with the  
requirements for the degree of Doctor of  
Philosophy

The University of Leeds

School of Languages, Cultures, and Societies

September 2016

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement

©2016 The University of Leeds and Jack J. Wilson

The right of Jack J. Wilson to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

## Acknowledgements

This thesis is dedicated to all those who I have neglected over the last four years.

I will also take this opportunity to specifically thank those, without whom, this thesis would probably not exist. First, I would like to thank the Arts and Humanities Research Council for funding the project. Second, I would also like to thank my supervisors, Bethan Davies and Catherine Davies. Without their patience and encouragement, particularly within the last few months, I'm not sure I would have managed to get to the end. Third, I want to thank Andrew Merrison, who is only connected to this thesis through a mutual obligation. However, Andrew's positivity was one of the few things that kept me going through the bleak periods. Fourth, Hannah Little, not just for being someone to talk to, but for also reminding me that I am not alone. Fifth, and most importantly, I want to thank Hazel. Not only has she been a constant source of support and encouragement throughout the PhD, but she has, on more than one occasion, reminded that there is more to life than academia.

## Abstract

Language and gesture are thought to be tightly interrelated and co-expressive behaviours (McNeill, 1992; 2005) that, when used in communication, are often referred to as composite signals/utterances (Clark, 1996; Enfield, 2009). Linguistic research has typically focussed on the structure of language, largely ignoring the effect gesture can have on the production and comprehension of utterances. In the linguistic literature, gesture is shoehorned into the communicative process rather than being an integral part of it (Wilson and Wharton, 2006; Wharton, 2009), which is at odds with the fact that gesture regularly plays a role that is directly connected to the semantic content of, in Gricean terms, “what is said” (Kendon, 2004; Grice, 1989). In order to explore these issues,

this thesis investigates the effect of manual gestures on interaction at several different points during production and comprehension, based on the Clarkian Action Ladder (Clark, 1996). It focusses on the top two levels of the ladder: Level 3 signaling and recognising and level 4 proposing and considering. In doing so, it explores gesture’s local effect on how utterances are composed and comprehended, but also its more global effect on the interactional structure and the goals of the participants. This is achieved through two experiments. The first experiment, the map task, is an interactive spatial description task and the second is an eye-tracked visual world task. These two experiments explore how gestures are composed during the map task, how gestures affect the real-time comprehension of utterances, and how gestures are embedded within the turn-by-turn nature of talk. This thesis builds a picture of the effect of gesture at each stage of the comprehension process, demonstrating that gesture needs to be incorporated fully into pragmatic models of communication.

# Contents

<b>1</b>	<b>Project overview</b>	<b>9</b>
1.1	Introduction . . . . .	9
<b>I</b>	<b>Literature Review</b>	<b>15</b>
<b>2</b>	<b>Pragmatics, gesture, and action</b>	<b>17</b>
2.1	The domain of pragmatics . . . . .	17
2.1.1	Introduction . . . . .	17
2.1.2	Types of meaning . . . . .	18
2.2	Gestures . . . . .	29
2.2.1	Gesture’s semiotic function . . . . .	29
2.2.2	Gesture’s relationship with the semantics of speech . . . . .	32
2.2.3	Gesture’s temporal relation with speech . . . . .	33
2.2.4	Gesture and Comprehension . . . . .	35
2.2.5	Gesture Production . . . . .	45
2.2.6	Gesture from the perspective of pragmatics . . . . .	55
2.3	Action is Laminated . . . . .	61
2.4	Summary & Research aims . . . . .	68
<b>II</b>	<b>Gesture, Meaning and Interaction</b>	<b>71</b>
<b>3</b>	<b>Methodology 1: The Map Task</b>	<b>73</b>
3.1	Introduction . . . . .	73
3.2	The Map Task . . . . .	79

3.2.1	This Map Task . . . . .	81
3.2.2	Set Up . . . . .	82
3.2.3	Maps . . . . .	82
3.3	Coding . . . . .	84
3.3.1	Projects and Sub-projects . . . . .	84
3.3.2	Move coding . . . . .	85
3.3.3	Semantic Coding . . . . .	87
3.3.4	Mechanical properties . . . . .	96
3.3.5	Coding Example . . . . .	96
3.3.6	Research Questions . . . . .	102
<b>4</b>	<b>An analysis of Gesture in the map task</b>	<b>105</b>
4.1	Introduction . . . . .	105
4.2	Production . . . . .	106
4.2.1	Gesture in the map task . . . . .	106
4.2.2	Participants and Move Types . . . . .	107
4.2.3	What kind of gestures are found in the map task? . . . . .	109
4.2.4	What is referred to (and in what medium)? . . . . .	111
4.2.5	Frames of Reference . . . . .	124
4.2.6	Perspective . . . . .	132
4.2.7	Motion . . . . .	140
4.2.8	Complexity of Semantic features in speech . . . . .	145
4.2.9	Gesture’s relationship to the word “around” . . . . .	146
4.3	Interaction . . . . .	153
4.3.1	Moves and gesture . . . . .	153
4.3.2	Sequential position of gesture . . . . .	160
4.4	Summary . . . . .	163
<b>III</b>	<b>Gesture and Comprehension</b>	<b>167</b>
<b>5</b>	<b>Methodology 2: The Visual World Paradigm</b>	<b>169</b>

5.1	Introduction . . . . .	169
5.1.1	This study . . . . .	172
5.2	The Visual World Paradigm . . . . .	173
5.2.1	A relevant history of the visual world paradigm . . . . .	174
5.2.2	Semantic Processing . . . . .	178
5.2.3	Informativity . . . . .	179
5.2.4	Visual World and Visual processing . . . . .	180
5.2.5	Visual world and gesture . . . . .	185
5.3	The current study . . . . .	190
5.3.1	Semantics of spatial descriptions . . . . .	191
5.3.2	Creating array items . . . . .	192
5.3.3	Utterances . . . . .	195
5.3.4	Recording and Stimuli creation . . . . .	198
5.3.5	Points of disambiguation . . . . .	199
5.3.6	Conditions and Variables . . . . .	201
5.3.7	Procedure . . . . .	201
5.3.8	Predictions . . . . .	202
<b>6</b>	<b>Analysis of Gesture in the Visual world</b>	<b>203</b>
6.1	Introduction . . . . .	203
6.2	Data Preparation . . . . .	207
6.3	Data Exploration . . . . .	207
6.4	Target Advantage . . . . .	209
6.5	Character Advantage . . . . .	223
6.5.1	Summary of Results for Character Advantage . . . . .	234
6.6	Competitor Advantage . . . . .	235
6.7	Response Time . . . . .	242
6.8	Discussion of results . . . . .	244

<b>IV Discussion and Conclusions</b>	<b>247</b>
<b>7 Discussion</b>	<b>249</b>
7.1 Speech and gesture in joint actions . . . . .	250
7.1.1 Signalling . . . . .	250
7.1.2 Recognising . . . . .	255
7.1.3 Proposing . . . . .	257
7.1.4 Considering . . . . .	258
7.1.5 Summary of main findings . . . . .	260
7.2 Gesture and Pragmatics: the predictability of gesture . . . . .	261
7.2.1 An alternative model of gesture production . . . . .	261
7.2.2 Predictability vs Tradeoff . . . . .	280
7.2.3 Predictability and the Hand-in-Hand hypothesis . . . . .	288
7.2.4 Predictability and the interface hypothesis . . . . .	291
7.2.5 Summary . . . . .	293
7.3 Limitations and Future Directions . . . . .	294
7.3.1 Map Task . . . . .	294
7.3.2 Visual World Study . . . . .	295
<b>8 Conclusions</b>	<b>297</b>



# Chapter 1

## Project overview

### 1.1 Introduction

Artists who produce depictions of human comportment have long recognised something that linguists are only beginning to take notice of. Edward Hopper's "A conference at night" (depicted in figure 1.1) is an excellent example of this fact.



Figure 1.1: Hopper's *A conference at night*

The painting depicts three individuals, two men and a woman, who are in a room with white walls and white columns. One of the men is stood, hands in pockets, wearing an overcoat and hat. It is unclear whether he has just arrived or is just about to

leave. The other man is sat on a desk, shirt sleeves rolled. The woman is also standing, partially obscured by the man in the foreground. The only light in the room appears to be coming from the window, however, closer inspection reveals that it is night outside. The light is presumably coming from a street light, giving some indication as to the height of the room. The question that scholars of human interaction may ask is: who is speaking? Hopper has left some clues. Visually, speaking is typically tied to the moving of the mouth, but none of the people in the painting appear to be moving theirs. What is noticeable is that the two standing individuals are both facing the sitting man; this is an indication they are attending to him. However, speakers also attend to their addressees. There is something more telling. The sitting man is holding his right hand in the air, gesticulating, producing what, in common parlance, is called a gesture. Many people looking at the painting may have a gut reaction that the sitting man was talking, without being able to define a single reason why this is likely to be the case. Others may immediately perceive the sitting man's raised hand as an indication that he is speaking. What is important to recognise is that Hopper knows that when people speak they regularly move their hands as well as their mouths. To quote similar sentiment from Erving Goffman (1964, p. 133), "[a] wagging tongue [...] proves to be one part of a complex human act whose meaning must be sought in the movement of the eyebrows and hand".

The painting is interesting for another reason. Hopper's image depicts three people interacting, but its title gives it extra meaning. The word "conference" changes the scene, imbuing the interaction with added meaning, but leaving enough mystery to make the viewer wonder what is actually happening in the scene. The word "conference" distinguishes their activity from other activities, such as a "meeting" or even a "conversation". Suddenly, their interaction has an air of seriousness. The fact that the image is enriched by the title is further emphasised by the phrase "at night". Looking at the painting, it is not at first obvious that the scene is happening at night. So now, the seriousness of the activity is elevated by the fact that it occurs at night, but it is also given a sense that the subject of the conference might be illicit.

The point of highlighting this painting is because it demonstrates an important type

of meaningful structure, that might be called semiotic unity (Enfield, 2009b; Enfield, 2013). Here, we understand that the image and the title are to be taken as one. The title enriches the meaning of the image, directing the viewer to salient features in the image and changing the meaning of others. This is because the title and the image have not been put together by accident, they were deliberately put together to convey a composite meaning (Enfield, 2009b; Clark, 1996). As observers we have no problem understanding that the two features—image and language—are to be taken together.

Therefore, there are two key features highlighted by Hopper’s work. The first is behavioural. When people talk they also move their hands. The second is semiotic. When placed together, signs are taken to form a meaningful structure. This brings us to the purpose of this thesis. The gestures that people produce when they talk are not merely movements, but are meaningful movements related to what a speaker is trying to communicate. Therefore, it is possible to analyse gesture and speech in a similar manner to painting and title (a fact observed by Enfield (2009b)). In fact, De Ruiter (2007) uses the metaphor of “postcards from the mind” to refer to utterances because, like postcards, they are a combination of text and image. To help emphasise this point, it is worth spending time looking at a famous example from the gesture literature.

One of the most renowned gesture scholars, David McNeill, has analysed gestures produced by people who are describing the events of Warner Brothers’ animations including the characters Sylvester and Tweety. The premise of the animation is that that Sylvester, a cat, wants to eat Tweety, a canary, but he never quite manages it. In one particular cartoon, Tweety is sat by a window and Sylvester is trying to climb a drainpipe. On his first attempt he climbs the outside of the pipe, but gets knocked down by Tweety. On his second attempt, Sylvester climbs up the inside of pipe before Tweety drops a bowling ball down the drainpipe, which pushes Sylvester down the pipe and down the street. It is this second attempt at climbing the pipe that forms the focus of the example.



Figure 1.2: Depiction of ‘rising hollowness’ (McNeill, 2005, p. 23)

The gesture depicted in figure 1.2 was produced while the speaker was saying: “and he goes up through the pipe this time”. The movement of the gesture was concurrent with the word “through”. McNeill describes the concept depicted in the gesture as RISING HOLLOWNESS, which accurately captures the motion of moving through a drainpipe. Importantly, rising hollowness is not something that can be directly expressed in English, but is a possible interpretation of “up through”. Therefore, it is possible to suggest that the gesture performs a similar function to the title of Hopper’s painting. The word “conference” affects the interpretation of the image and the rising hollowness gesture affects the interpretation of the words “up through”. The reason the two elements of Hopper’s painting and the two elements of the speaker in McNeill’s example are able to modify each other is because they are semiotically unified.

This thesis focusses on the semiotic unity of speech and gesture, investigating its production by individuals and how it is comprehended by addressees. The aim of the thesis is to explore gesture from the perspective of a particular subfield of linguistics, often referred to as lexical (Wilson, 2003) or truth conditional pragmatics (Recanati, 2010). Lexical pragmatics emphasises the role of intentions in the production and comprehension of utterances. And it is by focussing not just on the field of gesture research but to

the pragmatic enterprise more broadly, that this thesis offers a new contribution to both fields.

This thesis is divided into four parts. Part one explores the literature on pragmatics and gesture and tries to unify the hypotheses of both. Part two presents the map task (cf. Anderson et al., 1991; Anderson, 2006; Brown, 1995) methodology and analysis. The map task is a joint activity which must be completed collaboratively in order for participants to align on spatial descriptions. Therefore, it is the ideal methodology to elicit language and gesture, affording the exploration of their effect on interaction. Part three introduces and presents an experiment employing the visual world paradigm (cf. Huettig, Rommers, and Meyer, 2011), which uses eye gaze to explore the effect of speech and gesture on the real-time comprehension of utterances. Finally, part four summarises the findings of the experiments and proposes a new perspective on thinking about gesture within a pragmatic framework.

The next chapter begins by introducing the pragmatic framework.



# **Part I**

## **Literature Review**





# Chapter 2

## Pragmatics, gesture, and action

### 2.1 The domain of pragmatics

#### 2.1.1 Introduction

It is often remarked by gesture scholars that meaning does not start with words (cf. McNeill, 2015; Kendon, 2004; Enfield, 2009b; Clark, 1996). The suggestion is that meaning is a product of speech and gesture (amongst other things) and to focus on linguistic meaning severely underplays the communicative and cognitive importance of gesture. Therefore, the notion of utterance, the typical unit of linguistic production, should never be considered to be monomodal. From this point on, I will adopt the convention of referring to those who produce utterances as utterance producers (or just producers) and those who comprehend utterances as utterance comprehenders (or just comprehenders). Utterances are multimodal (we will return to this in section 2.2 below). However, it has also been recognised in linguistic pragmatics, even though the focus has mainly been on the spoken linguistic component of utterances, that meaning does not start with utterances, but behaviours (Sperber and Wilson, 1986). However, in this thesis I do not adopt a single pragmatic theory. The purpose here is not to champion one pragmatic theory over another, but to use key ideas within the pragmatics literature to explore gesture. The opening of this chapter will focus on how pragmatics deals with (predominantly linguistic) utterances.

Language, in the words of Lewis (1969), is a solution to an everyday coordination

problem. Language developed to offer a shortcut to the problem of coordinating actions in the world and representations of the world. However, language is not a complete solution, there is a problem of viability (cf. Carston, 2002), because language does not (and perhaps cannot) represent everything we are capable of thinking about. What's more, like every shortcut there is still a degree of travelling required to get from A to B, which in this case represents the difference between what an individual encodes linguistically and what they intended to communicate. Linguistic pragmatics can be thought of as the study of the distance that still remains to be covered once the shortcut of language is taken.

### 2.1.2 Types of meaning

Much of the Post-Saussurean linguistics in the twentieth century is built upon the idea that a sign has meaning because it specifies a standing-for relationship between a certain signifier and signified (Kockelman, 2005). For example, a word stands-for the thing it refers to, which is why the word “plank” typically refers to a long, flat and thin piece of wood. Another cornerstone of modern linguistic theorising rests on the idea that certain token behaviours have the ability to refer to things because they are instances of types of behaviour available to individuals abstractly. Thus we are able to use tokens based on their abstract type. For example, a tokening of the word “plank” will typically differ from moment to moment, due to (amongst other things) the constraints of the biological apparatus involved in its articulation. However, the token is typically understood to be a tokening of the type of behaviour which may be used to refer to a long, flat and thin piece of wood. This relationship between a word and what it typically means is a semantic one, since the word has a meaning because of conventions surrounding its use. Moreover, those who use a word can be fairly assured that the meaning of that word will be within the common ground of their speech community. Therefore the word allows members of the speech community to coordinate linguistically (cf. Clark, 1996; Lewis, 1969; Schelling, 1960). However, this relationship between a token and its type is not as rigid as it often seems. In fact people frequently use token behaviours, not to refer to what they typically encode, but to evoke some of the elements typically related with a

token's type. Imagine an utterance of:

(2.1) John is a plank

Upon hearing (2.1) it is not typically assumed that the producer is suggesting that John is a long, flat and thin piece of wood, but rather he has certain properties typically associated with one—density, for example. With this in mind, pragmatics can be thought of as the study of how context guides interactants, providing these apparently rigid, conventionalised type-token relations with almost boundless flexibility (cf. Levinson, 1983).

This raises a question: if the comprehension of an utterance like the one in (2.1) is not always the result of convention, then what process guides a comprehender to understand its token specific meaning? The view espoused in much of the pragmatics literature is that people do not communicate meaning directly, but rather they present behaviours from which their intentions can be inferred<sup>1</sup>. As Sperber (1995, p. 191) notes the “only thing that is ever produced by one person for another person to see or hear is behaviour and the traces it leaves behind”. Understanding the nature of these traces is key to understanding communication generally. In most post-Gricean frameworks, the intentions involved in communication can be unpacked as (1) an intention to inform someone of something, and (2) an intention to make that intention manifest by producing some behaviour (out of the set of all possible behaviours available to them) from which a comprehender could reliably infer the first intention (Grice, 1957). These two intentions have been referred to as the *informative* and *communicative* intentions respectively (Sperber and Wilson, 1986).

By inferring these intentions, a comprehender can determine the meaning a communicator intended to convey—a process frequently referred to as *mind-reading* (Bara, 2011; Sperber and Wilson, 2002). Mind-reading is the product of the incredibly attuned human ability to represent someone else's thoughts, often called Theory of Mind (cf. Frith and Frith, 2005). Moreover, it is this mind-reading ability, and not conventions, that are the true driving force of communication. From the producer's perspective, utterances are recipient designed (Garfinkel, 1967). They are produced according to expectations based on the predicted reaction of a comprehender (Kockelman, 2012). In other words, A pro-

<sup>1</sup>See, amongst many others, Bara (2011), Grice (1975), Levinson (1983), and Sperber and Wilson (1986).

duced behaviour  $p$ , predicting, based on  $A$ 's knowledge of  $B$ , how  $B$  will react. Further,  $B$  will represent  $A$ 's behaviour as an attempt to inform  $B$  of  $i$  (which is  $A$ 's informative intention).  $B$  reacts to  $A$ 's behaviour inferring  $A$ 's prediction (i.e., that  $A$  believes  $p$  will allow  $B$  to infer  $i$ ). It is because the conventions associated with lexical items reduce the entropy associated with the relationship of  $i$  and  $p$ , that they act as an incredibly stable communicative resource. Conventions increase the reliability of communicative inference.

Furthermore, in terms of using language, Grice (1957; 1975) made a distinction, echoing the type-token relation, between **WHAT IS SAID** (what is encoded by the language) and **WHAT IS MEANT** (what the utterance producer is taken to mean on a particular occasion). Typically, what people actually mean is not exhausted by what they say, and hence the speaker's intentions are not identical to the conventional meaning of the language. One of the central debates within linguistic pragmatics is the role that inference plays in understanding *what is said*. Here, I adopt the perspective that all linguistic content is typically modulated to fit the context in which it is comprehended. The process is referred to as a contextualisation (Sperber and Wilson, 1986). In this sense we can think of pragmatics as the study of the way in which communicators use and comprehend disparate sets of phenomena (e.g., the ambient environment, the behaviours of another individual, which may be conventional or non-conventional, memories, feelings etc.) bringing them together to form an understanding of what is being meant during interaction. This is as true for producers as it is for comprehenders. For example, imagine Anne, having been at work all day, comes home and is engaged by Bob in the following interaction <sup>2</sup>:

(2.2) **Bob** How was your day?

**Anne** God, I need a drink

+> Not Good

In (2.2) what Anne explicitly says (“God, I need a drink”), is not the same as what she means to communicate (“Not good”). Since we (and Bob) can assume that Anne is per-

---

<sup>2</sup>Here, the symbol + > is used to denote the implied meaning of an utterance.

fectly capable of producing an utterance regarding the quality of her day (e.g., “Not good”) but doesn’t, we can infer that she is guiding Bob to the conclusion that she is doing something more. We can describe what Anne means as the implied meaning or *IMPLICATURE* of her utterance (cf. Grice, 1975; Levinson, 2000; Recanati, 2004; Sperber and Wilson, 1986). If pragmatic processes are involved in arriving at the implicature then it seems that what these processes do is build on what is provided by semantics, a view which may be thought of as the classic semantics/pragmatics divide (cf. Gazdar, 1979). However, this view misses an integral step in the inferential process because the literal (conventional) meaning associated with Anne’s utterance does not warrant Bob arriving at the conclusion that her day was not good—Anne might have just come home after going out for a run, for example. In which case her utterance could be understood as relating to her desire to avoid dehydration. Therefore, we might suggest that the inferential processes involved in implicature derivation must be working on the meaning of the individual sentential constituents of her expression, suggesting that understanding *WHAT IS SAID* and *WHAT IS MEANT* are mutually adjusted to fit each other (cf. Carston, 2002; Recanati, 2004; Recanati, 2010). This has been taken to suggest that there is perhaps a third level in between *WHAT IS SAID* and the implicature, sometimes labelled a generalised conversational implicature (GCI) (cf. Grice, 1975; Levinson, 2000). A GCI is an implicature that is carried out unless there is evidence to suggest the producer means something else. Here, however, the term *EXPLICATURE* (Sperber and Wilson, 1986) is preferred since it rests on the idea that there often isn’t enough information derivable from *WHAT IS SAID* for it to warrant its own stage in the model of comprehension.

Explicature can be used to refer to both what Anne has actually said and any contextual enrichment or modulation required for the expression to express a truth evaluable proposition, but it stops short of deriving an implicature. For example 2.2, in order to arrive at the explicature one would have to resolve the pronoun “I” which would have a different interpretation if the utterance was produced by another speaker. Furthermore, the word “drink” does not contribute to the understanding that she has not had a great day, but can only do so if its meaning is narrowed to mean \**DRINK* (viz. *ALCOHOLIC DRINK*). This further highlights the fact that the processes involved in implicature

and explicature derivations are mutual since the motivation behind the narrower understanding of “drink” is due to the expectancy that Anne’s answer to Bob’s question will be about the quality of her day. Discourse particles, such as “God”, are also important since they help guide the comprehension process. In this case it provides information about the utterance as a negative assessment. Such elements have been described as conveying not expressive but “procedural” meaning (Blakemore, 2002). The picture we are left with is a model of language comprehension, levelled functionally but not temporally, in which both the explicature and implicature are mutually adjusted in order for a token specific meaning to be derived. Ultimately, what this example highlights is that the process of arriving at a token meaning through convention alone seriously underdetermines WHAT IS MEANT and inferential processes are involved in comprehension from the bottom up. This idea has been argued for in many guises in both the philosophically oriented (Carston, 2008; Recanati, 2004; Recanati, 2010) and computationally oriented literature (Blutner and Zeevat, 2004; Blutner, de Hoop, and Hendriks, 2006; Parikh, 2010).

However, we have so far only considered how words may form an utterance. As stated above, communicative behaviours are often more complex and include gestural elements. Such gestural elements are of paramount importance to pragmatic theories of meaning because arguably there aren’t any conventions upon which enrichments and/or modulations can work—there seems to be no semantics (in the traditional linguistic sense) of gesture. For example:

(2.3) **Bob** How was your day?

**Anne** God, I need a drink

((mimetically shapes her hand into a gun shape, puts her index and middle finger (representing the barrel of the gun) to her head and pulls her thumb from an outstretched position into her hand (representing the hammer of the gun)))

+> Not Good

In this case, the gestural element seems to be guiding the comprehension of the linguistic elements and thus are not strictly part of the expressive meaning, and hence not part of

the explicature. In other words, Anne’s mime highlights her negative attitude. However, consider:

(2.4) **Bob** How was your day?

**Anne** ((mimetically shapes her hand into a gun shape, puts her index and middle finger (representing the barrel of the gun) to her head and pulls her thumb from an outstretched position into her hand (representing the hammer of the gun)))

+> Not Good

In this case, Anne’s gesture is the only potentially communicative element of Anne’s behaviour and is all an addressee has to go on. Therefore, it is unclear how meaning can be derived from the behaviour since it does not really encode anything; it is not a token of a type but a one off token—what Kockelman (2005) has referred to as a *singularity*. So far it has been assumed that semantic processes work on a base level providing pragmatic processes with something to ‘work on’. However, in this example there is nothing for semantic processes to work on—Anne’s behaviour is not conventional<sup>3</sup>. Therefore, from a traditional perspective, in this example the onus of understanding Anne’s communicative behaviour falls exclusively on pragmatic processes. This suggests that understanding the meaning of Anne’s behaviour in example 2.4 is not linguistic. However, as we will see below, for many examples of gesture it is not so easy to suggest that they are non-linguistic.

One final element, which isn’t part of Anne’s behaviour, but is important for understanding what she means, is context. Context is an elusive thing to define. It may be thought of as everything that is necessary for understanding the meaning of an utterance, but is not part of that utterance<sup>4</sup>. For example, Anne may have told Bob earlier that she was dreading her day because she had not finished an important piece of work and didn’t want to face her boss. This interaction could lead Bob to the assumption that

<sup>3</sup>One might object at this point and argue that Anne’s gestural contribution is conventional. I will concede that Anne’s gesture is certainly not novel, but it has been used for the purpose of explication. However, I would argue that an individual with no prior experience of a gesture similar to Anne’s would have a better chance of gleaning what she means than someone who has had no prior experience with the English language would have of comprehending the meaning of what she said.

<sup>4</sup>See Levinson (1983) for an exploration into the nature of context; and see Sperber and Wilson (1986) for a cognitive approach to context.

Anne was not looking forward to her day and thus likely not to enjoy it, which might have made the conclusion of “not good” more salient within that context. Here, I adopt the relevance theoretic notion that context is not something that exists “out there” in the world, but follow the relevance theoretic idea that context is chosen during comprehension (Sperber and Wilson, 1986). This view is also highlighted by Kockelman (2012).

In summary, pragmatics, in linguistic terms, can be thought of as meaning minus truth conditions (Gazdar, 1979) or context dependent meaning (cf. Recanati, 2010, for a review). However, such linguistic views are typically built on a conception of language as fundamentally different from other human action and meaning in language as different from meaning elsewhere in our everyday lives. Furthermore, there is often a suggestion of specialised cognitive machinery for dealing with language (cf. Sperber and Wilson, 1995). Since the primary interest here is in how non-linguistic behaviours are part of communication, then such linguistic-centred theories are often not useful at best, and harmful at worst. An alternative approach taken by Enfield (2009b) is to view the production and comprehension of linguistic behaviour as analogous to any other form of decision making, and as such is subject to heuristics. The decisions to be made during communication revolve around the behaviour-producer deciding what behaviour to produce and the behaviour-comprehender deciding why that behaviour was chosen. As stated earlier, conventional linguistic behaviours can be thought of as shortcuts, or community heuristics, easing the burden involved in making such decisions. From this perspective, speech and gesture are different methods of easing the burden. Speech (or more specifically, language) relies on a foundation of conventional correspondances between sign and meaning, whereas gesture relies on similarity or contiguity between sign and meaning. In this sense, meaning in communication is encoded or non-encoded (Enfield, 2013), what Peirce (Peirce, 1955) called *symbolic* and *indexical* or *iconic* respectively.

In the framework of *fast and frugal heuristics* (Gigerenzer, Todd, and ABC Research Group, 1999) we can think of a decision making process as a three stage process involving: (i) locking on to the target decision to made; (ii) a search for and narrowing of candidate alternatives; (iii) locking off once a decision has been made. In terms of language comprehension, Enfield (2009b) has described these as the (i) on-switch, (ii) search, and



(iii) off-switch respectively. Following Enfield's approach we can describe these three stages as follows:

#### ON-SWITCH

What individuals lock onto in order to begin their search for meaning is fairly uncontentionally believed to be the intentional nature of communicative behaviours described above. In other words, if a comprehender believes that a behaviour is intentionally communicative then it automatically requires attention. It has been demonstrated that such intentions are not just the basis of individual communicative processes but are also a key factor for language acquisition generally (cf. Behne, Carpenter, and Tomasello, 2005; Gergely, Bekkering, and Király, 2002; Tomasello et al., 2005). Communicative intentions may be made apparent through eye-gaze or stance (Hanna and Brennan, 2007; Mol et al., 2011; Neider et al., 2010; Richardson, Dale, and Tomlinson, 2009). Conventional linguistic behaviours, by their very nature, are specialised behaviours indexing communicative intentions, and thus bring with them a presumption that they are worth processing (Sperber and Wilson, 1986).

From the producer's perspective, the on-switch is the problem that needs solving. This problem could be getting someone to know something (e.g., that it's raining) or it could be getting someone to do something (e.g., closing the window).

#### SEARCH

The search stage involves the process of constraining which elements<sup>5</sup> are worth including in the comprehension process and various notions have been developed to describe this process. These include notions of (bounded) rationality, informativity, relevance, propriety, salience, mutuality, amongst many others (cf. Clark, 1996; Goldstein and Gigerenzer, 2002; Grice, 1975; Sperber and Wilson, 1986). It should be emphasised that communicative behaviours do not mean anything without some interaction with context, but have what may be referred to as 'meaning potential' (Recanati, 2004). Fur-

<sup>5</sup>The number of elements that can play a role in the search stage are potentially boundless. Obvious elements are speech and gesture, but search may be constrained by the clothes someone is wearing or the posture they are adopting. Further, search will also be constrained by context, which may or may not be intentionally invoked by the utterance producer.

thermore, while a communicative behaviour is evanescent, context may inexhaustibly provide reasons to keep the search ongoing. In some cases, even long after a sufficient understanding has been selected, people may continue to derive further implications that are not strictly necessary for understanding the utterer's intentions on a specific occasion<sup>6</sup>.

From the producer's perspective, this will look quite similar. The producer must use their knowledge of the comprehender and their knowledge of language/behaviour something that will achieve their goal, which acted as the on-switch to this process.

#### OFF-SWITCH

Finally, we have the question of why people stop the process of comprehending, what Enfield, 2009b, p. 226 calls the sixty-four-thousand-dollar question. Considering the fact that human processing capabilities are not infinite, this is a question that certainly deserves an answer cf. Gigerenzer, Todd, and ABC Research Group, 1999, pp. 5–15. However, few have offered well substantiated suggestions. Candidate explanations have included the idea that processing stops once expectations of relevance have been met (Sperber and Wilson, 1986) or once some communicative behaviour has been understood sufficiently enough to elicit an appropriate response—what is often phrased 'sufficient for current purposes' (cf. Clark, 1996, p. 222). However, such theories often present more question than answers. For example, in his review of Sperber and Wilson's (1987) relevance theory, Clark asks "relevant to what?" (see pages 714f.) Here, I am not so interested in why people stop processing, but I take it for granted that they do. My concern is predominately with what people include as constraining factors during the search for meaning. And therefore, what communicative behaviours should be included in a model of communication.

Once again, this stage can be represented as being quite similar for producer and comprehender. While the comprehender stops once they believe the producer's goal is satisfied, the producer stops once they believe their goal is satisfied. While the evidence

---

<sup>6</sup>For example, there are countless examples from my childhood where I thought, at the time, that I had perfectly understood what an adult (parent or teacher, for example) had meant, only to realise years later that I did not have access to some contextual information. Disney Pixar have perfected the art of pitching the same information at two levels, one for the children in the audience and the other for the adults.

that comprehender and producer take as the satisfaction of the goal may differ, sufficient coordination is usually achieved. If it wasn't, then communication would rarely work.

This model can be conceptualised in relation to the classic Peircean (1955) description of the third. A third can be thought of as the smallest unit of interest to scientific endeavour. Peirce's most well known formulation of the third involves the sign, object, and interpretant. Peirce's work has received much attention in the twentieth century (for an excellent overview of Peirce's work see Short (2007)). However, in this thesis, it is only necessary to briefly outline the three-part nature of signification. For Peirce, a sign stands for something (variably referred to as its referent, denotation, or extension), which Peirce calls an object. For example, the word "dog" conventionally refers to four-legged, domesticated canines. However, if one studies the relationship between a sign and an object exclusively, then it is only possible to describe the abstract or potential. We can say things like, "the use of the sign "dog" conventionally refers to four-legged domesticated canines" or instruct someone by saying "if you use the sign "dog" someone will understand you to be referring to four-legged domesticated canines". However, we cannot say how an actual instance of the sign "dog" was significant. The two-part model for analysing the relationship between sign and object, which (Kockelman, 2005) calls a relationship of *standing-for*, must be enriched with a third part. This third part is what Peirce called an interpretant. The interpretant is whatever follows from the production of the sign, insofar as the sign was taken as being related to its object. From this perspective, the relationship of a sign standing-for an object is secondary to the relationship of correspondance between sign and interpretant, and it changes the notion of object from being something that is objective (e.g., four-legged domesticated canine) to something which Kockelman (2005, p. 242) calls a *correspondance-preserving projection*. An example can help clarify this distinction. Reusing example 2.2 above (presented here as example 2.5):

- (2.5) **Bob** How was your day?  
**Anne** God, I need a drink  
 +> Not Good

If we take Bob's utterance as a sign, then we can take Anne's utterance as an interpretant. From this perspective, the object is the correspondance between Anne's response and Bob's utterance. That is, whatever makes the sequence of behaviours represented as Bob and Anne's utterances meaningful is the object of Bob's utterance. According to this understanding of example 2.5 both sign and interpretant are behaviours, but this is not necessary. For example, if we took Anne's utterance as a sign, then it is possible to argue that the implied meaning "not good", is its interpretant. However, "not good" is not produced as a behaviour but is represented in Bob's mind. In this case, the object would have to be Anne's intention in producing the utterance, because it is Anne's purpose that acts as a correspondance preserving projection between the utterance "God, I need a drink" and the implicature "not good". Once the three-parted nature of meaning is acknowledged, it becomes possible to model the meaning of any human behaviour, either externally as a sequence of behaviours, or mentalistically as intentional states and behaviours. From this perspective, it is easy to analyse Anne's behaviour in example 2.4 where she did not produce a linguistic utterance, but produced a mime instead. Anne's gesture can either be taken as an interpretant of Bob's utterance or it can be taken as a sign which has the potential to stand-for her communicative intention and thus allow the inference of "not good". Relating this back to fast and frugal heuristics, it is possible to suggest that for any decision making process, the on-switch is the sign, the search is the object, and the off-switch is the interpretant. Furthermore, this suggests that both the production and comprehension of utterances can be modelled as decision making processes.

So far in this chapter, we have only dealt with one gesture (see examples 2.3 and 2.4), which seems to be quite different from McNeill's example discussed in Chapter 1. In McNeill's example the gesture and the speech seemed to be about the same thing (i.e., the way Sylvester climbed the pipe) whereas in example 2.4, the mime of the gun was a reflection of Anne's attitude to what she was saying. If we are to understand the nature of gesture, then it is crucial not just to have a model of meaning, but to understand what types of gesture one might find, and how they are produced and comprehended. The next section presents an overview of the literature on gesture.

## 2.2 Gestures

Ekman and Friesen's (1969) classification system of gesture set the scene for much of the work on gesture in the twentieth century. However, their work has been criticised by Kendon (2004) because their classifications "have not been established according to a common set of criteria" and that "members of one category are also members of another category, depending upon the point of view of the analyst" (p. 97). Here, following much of Kendon's work<sup>7</sup>, in which speech and gesture are fundamentally part of an utterance, we may argue that it is not useful to define gestures according to their structural properties, but should instead define them according to some external categorisation under which both language and gesture can be explored. Candidate categories include semi-otic ground, temporal organisation, or referential/semantic meaning. Each of these will be covered in turn.

### 2.2.1 Gesture's semiotic function

Enfield (2009b, p. 18) has developed a semiotic taxonomy of gesture that builds on the classic, Peircean categories (Peirce, 1998). Peirce describes three distinct relationships between signs (e.g., a word) and objects (e.g., the thing the word refers to). This relationship is sometimes called a sign's ground. These are: (i) *iconic*, where the sign contains perceptual qualities similar to some object; (ii) *indexical*, such that a sign exists in a contiguous relationship with an object (e.g., spatial, temporal, or causal); and (iii) *symbolic*, in which the relationship is the result of some predetermined convention or contact (cf. Peirce, 1998). Enfield's taxonomy acknowledges that a single sign often includes multiple grounds (this will be covered in more detail below). Enfield's taxonomy (2009, p. 18) is as follows:

---

<sup>7</sup>See also, amongst many others, Armstrong, Stokoe, and Wilcox (1995), Kelly, Özyürek, and Maris (2010), and McNeill (2005).

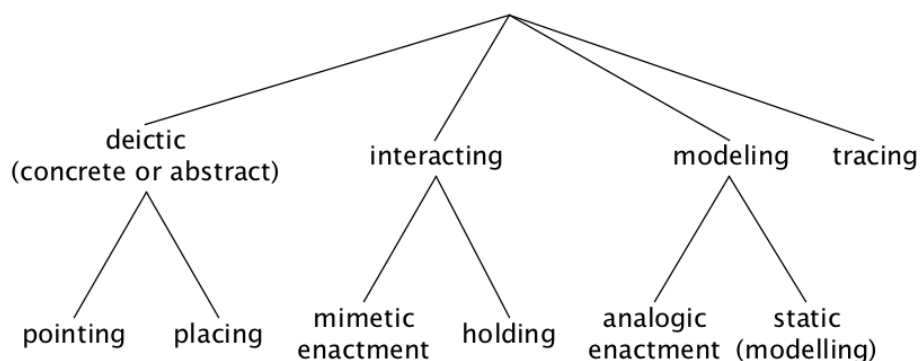


Figure 2.1: Enfield’s (2009, p. 18) classification of semiotic devices

Enfield further classifies these different gestures according to their semiotic functions as follows.

#### DEICTIC

Deictic gestures can be divided into “concrete”, referring to something in the ‘real’ world and “abstract”, referring to something in gesture space (Melinger and Levelt, 2004). Regardless of whether or not deictic gestures are concrete or abstract, they can be placed into two groups: pointing and placing (cf. Clark, 2003). Placings are arguably just as common as pointings but have received far less attention in linguistics. The semiotic relationship in both of them is indexical, however pointings involve the creation of a vector from sign to object, whereas placings highlight the spatial placement of an object thus making it salient for some observer. For example, if an individual wanted to direct another person’s attention to something then they might point to that thing or they might pick that thing up and place it for them to see. (Of course the ability to use placing gestures is dependent on the thing being indicated—while it may be possible to place a bottle of shampoo, it would not be so easy to place a sports car or a jumbo jet.) Thus the purpose of deictic gestures is to engender joint attention regarding some object (Bangerter, 2004). Deictic gestures, along with joint attention, have been argued to be key foundations of human communicative abilities both in terms of phylogeny (Hewes, 1981; Kendon, 1991) and ontogeny (Grassmann and Tomasello, 2010; Tomasello, Car-

penter, and Liszkowski, 2007).

#### INTERACTING

Interacting gestures come in two forms: *Mimetic enactment*—which is similar to *pantomime* (cf. McNeill, 1992) or what Ekman and Friesen (1969) called *kinetographs*—and  *Holding*. An example of mimetic enactment would be holding both hands out parallel as if holding a pole and then repetitively twisting one hand referring to the action used when accelerating on a motorbike, thus referring to either the motorbike or the act of riding one (Anne’s gesture, see example 2.3 above, could be thought as mimetic enactment).  *Holding* is similar but the action is simply one in which the referent is being held. For example, imagine the action just described minus the hand movement, which might be used to refer to how one might hold a pole while water-skiing. These gestures are iconic, since the activities referred to by the gestures are linked perceptibly (i.e., the gestures look like the referent actions). However, they are also indexical because, in the case of holding gestures, they do not represent the referent itself, but rather direct the recipient to salient features of the comportment of an individual interacting with the referent. And, in the case of mimetic enactment, an individual only produces part of the prototypical behaviour associated with the referent.

#### MODELLING

*Modelling* is divided into two categories: *analogic enactment* and *static modelling*. During analogic enactment the gesture mimics the movement of some referent. For example, an undulating hand movement in front of and across the body might depict the movement of a snake. Alternatively, for static modelling the hand takes on the form of the referent, such as a clenched fist being used to represent the heart. These gestures are iconic. They may also be indexical since the hand may not represent the entire object but use a model from which the whole object can be inferred. For example when modelling a snake it would be impossible to model every aspect, such as the scaled skin, and thus these aspects must be filled in during comprehension.

## TRACING

Tracing is one of the most interesting gesture types because it involves the hands, most often the fingers, drawing in abstract gesture space. Importantly, it is not simply the movements that have meaning but the space in front of the individual becomes a semiotic platform, without which the gesture would not have meaning. Amazingly, people frequently use these gestures without any prior agreement that they are going to do so. These gestures are indexical and iconic.

Tracing and modelling are linked in several respects. They are both gestural constructions representing the spatio-visual properties of some non-present referent. They are often grouped under the heading *representational gestures* (cf. Holler and Beattie, 2003), however, I believe distinguishing them can play a pertinent role when discussing the gestural contributions to communication. Furthermore, we may also add emblematic gestures to this taxonomy. Gestures such as the “ok” gesture are considered emblematic (Ekman and Friesen, 1969, pp. 63–92), but these can be thought of as more like gestural equivalents to words, and therefore the semiotic ground invoked by them are predominantly symbolic.

### 2.2.2 Gesture’s relationship with the semantics of speech

If we are to understand how speech and gesture are packaged to form a single unit, then it is important to understand how such gestures collocate with the spoken components. For this, the semiotic relationships invoked by a gesture are not particularly useful because they are not inherently linked to the specifics of what a communicator is attempting to refer to. For example, a pointing gesture may refer to an object in one situation, whereas a modelling gesture may be used in another. Kendon (2004, pp. 176f.) provides six different ways in which gestures seem to collocate with the semantic/referential meaning of the language of an utterance:

- (1) There are gestures which have a ‘narrow gloss’ used in parallel with referentially or semantically equivalent words or phrases. Such gestures appear to be completely synonymous with speech and thus the semantic relationship seems to be one of complete redundancy. For example, using an emblematic gesture



such as the “ok” gesture whilst providing the verbal expression “ok”.

- (2) Gestures with a ‘narrow gloss’ are not always completely semantically redundant, however. In such cases gestures convey additional semantic information not present in the speech. For example, using an emblematic gesture such as the “ok” gesture whilst providing the verbal expression “I’m not too bad”.
- (3) Gestures may provide information which is more specific than that conveyed in speech alone. Kendon suggests that these gestures frequently occur when describing activities (p. 185). For example, if someone is describing a throwing action the gesture might be an enactment of elements which relate to the type of thing being thrown.
- (4) In a similar way, gesture may be used to create the representation of an object of some kind. They may present an exhibit or specimen (p. 190) such as presenting a clenched fist whilst producing the word “heart”.
- (5) Gestures may be used to specify the shape, size and spatial characteristics of a single object or the relationships between objects. For example, tracing out a box while saying the word “square” or forcibly bring one’s hands together while describing a crash.
- (6) Gestures can be employed to create objects of reference for deictic expressions, which, as stated above, may be abstract or concrete. For example, placing hands some distance apart and saying “it was this big”, in which the referent of the expression is conveyed through the gesture.

### 2.2.3 Gesture’s temporal relation with speech

This co-referential structure of gesture and speech is also carefully timed so as to form a single temporally bounded *gesture unit* (Kendon, 2004). The *gesture unit* is the totality of visual activity bookended by two *rests* or *home* positions (Sacks and Schegloff, 2002), or moments of relaxation during which the articulators are not being employed productively. Regarding the hands, the canonical rest can be thought of as being ‘on the lap’ or ‘on a table’. Typically, *rest* and *home* are seen as being synonymous, however here I will contend that *home* is relative to the current talk and that a unit may be bound by physically different home states. For this reason, I will make the following distinction

between the two terms: (i) *Rest* refers to the point at which the communicator is at ease or no longer actively engaged in the communicative activity; they may, however, still be engaged in some broader activity (e.g., putting furniture together); and (ii) *Home* can be thought of as referring to some relative neutral point from which a communicator may pick up their next utterance or turn; it is important for the definition of home that the positioning of the hand(s) does not seem to deter the other communicator(s) from making their contributions.

The gesture unit represents the entire movement excursion from home and then back again, and is composed of three to five (because holds are optional) gesture phases (Kendon, 2004, pp. 113–124). These phases are:

1. The *preparation phase* consisting of the incipient stages of the gesture. It represents the initial movement away from *home*.
2. The (optional) *pre-stroke hold*, during which the hand is held in its position at the end of *preparation phase*. This phase allows the utterance producer to pause in order to increase semantic synchronisation between speech and gesture.
3. The *stroke* which may be thought of as the nucleus of the gesture and is typically the most meaningful phase.
4. The (optional) *post-stroke hold*, during which the hand is held in its final position. It was originally observed by Kita (1993) and it allows the gesture producer to elongate the gesture’s composition, often fitting it to the spoken elements of an utterance.
5. Lastly, the *recovery* is the movement back to *home*.

The first four of these phases combined are referred to as the gesture phrase (see figure 2.2). The gesture phrase is typically the meaning bearing element of the gesture.

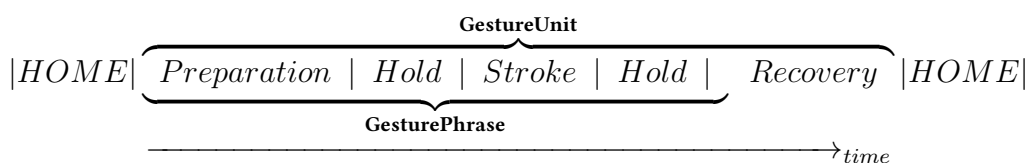


Figure 2.2: composition of a gesture unit

Kendon argues that gesture phrases closely collocate with the tone units (cf. Crystal and Davy, 1969, pp. 24–40) of the accompanying speech. He further suggests that:

Tone units are packages of speech production identified by prosodic features which correspond to units of discourse meaning. In the same way, gesture phrases are units of visible bodily action identified by kinesic features which correspond to meaningful units of action such as pointing, a depiction, a pantomime or the enactment of a conventionalized gesture.

(Kendon, 2004, p. 108)

Therefore, tone units and gesture phrases are most accurately conceptualised as fundamental partners in a unit of interactional behaviour, all components of which are part of the utterer's final product (p. 134).

If we accept the idea that speech and gesture are synchronised both temporally and semantically, and that they form complex semiotic structures, then we can begin to ask what sorts of information are presented through gesture and whether they have an observable effect on communication. Evidence comes from a variety of sources. The rest of this section will be dedicated to exploring the comprehension and production of gesture (comprehension of gesture is explored empirically in chapters 5 and 6 and the production of gesture is explored in chapters 3 and 4).

### 2.2.4 Gesture and Comprehension

The central question reviewed in this section is whether gesture has an effect on the comprehension of utterances. In early gesture research there was a common assumption that gestures were communicative (cf. Kendon, 1994, for a review). However, much of the early evidence that gesture serves a communicative function came from observational analyses. Krauss, who once disagreed with the communicative nature of gesture (Krauss et al., 1999), provided experimental evidence against the view that gestures are communicative. Instead, it was found that gesture served little function beyond the communicative function already present in the concurrent speech (Krauss, Morrel-Samuels, and Colasante, 1991). In the last twenty years there has been a surge in experimental

studies exploring the effect of gesture on comprehension. This section outlines the major findings, first exploring behavioural studies and then neuropsychological ones.

### **Behavioural Studies**

Beattie and colleagues (cf. Beattie and Shovelton, 1999b; Beattie and Shovelton, 1999a; Beattie and Shovelton, 2001; Beattie and Shovelton, 2002) set out to directly question Krauss's position on the communicative nature of gesture. They used a methodology that employs structured interviews and questionnaires to explore whether or not comprehenders receive information through gesture. In order to do this, they recorded encoders narrating stories from comic strips and then interviewed respondents, asking questions regarding specific semantic features in the narrations. Semantic features included the shape, position, and size of objects, and the direction, orientation, manner, and speed of actions. This methodology has subsequently become known as a semantic feature analysis (Gerwing and Allison, 2009b), because rather than focus on the form of speech and gesture it focusses on the possible properties of a referent speech or gesture might be highlighting. In the interviews, respondents were asked questions specifically designed to investigate whether participants would recall information depicted in gesture. This method, therefore, affords the analysis of whether or not a particular semantic feature was recalled as a result of that feature being presented through speech and/or gesture. This is a useful methodology because as Beattie and Shovelton (1999b, p. 442) argue, it allows one to study the relationship between gesture and the “world out there, the world waiting to be encoded into speech or gesture, or both”.

In Beattie and Shovelton (1999a), they asked a series of yes/no questions that related to specific semantic features described in the videos of the narrators. Respondents were either shown the video or heard the audio on its own. Beattie and Shovelton (1999a) found that gesture correlated with an increase in accurate answers, but also that it was only gestures pertaining to relative size and relative position of objects that seemed to significantly communicate information about an object above and beyond the information contained in speech. In Beattie and Shovelton (1999b), rather than using yes/no questions, a series of structured interviews were used. Respondents were also asked

how confident they were with their answers. In this study, they also included a vision only condition, in which respondents saw the video with muted audio. The results for this study were similar to Beattie and Shovelton (1999a), in that respondents were more accurate when they saw gesture and heard speech and they seemed to extract more information relating to shape and relative position of objects. Additionally, in this study Beattie and Shovelton (1999b) found that respondents were more confident in the vision only condition when questions concerned the relative position and shape of objects. These studies provide strong evidence that gesture is performing a communicative function.

Employing a distinct methodology, Kelly et al. (1999) explored the effect of gesture on enriched utterance comprehension (e.g., understanding indirect requests). They showed that pointing gestures are influential for comprehending statements as indirect requests. Furthermore, when comprehended together, speech and deictic gesture are better together than either speech or deictic gesture for observers comprehending a statement as an indirect request. In another experiment, they demonstrated that participants were better at comprehending the indirect object of an utterance when speech and gesture occur together, rather than when just gesture occurred. And, in a final experiment, they explored the use of iconic gestures, showing that, not only did gesture facilitate the retrieval of information presented in speech, but information presented through iconic gestures was recalled as part of the spoken component of the utterance. This study, therefore, suggests that not only are gestures comprehended, but that comprehenders are potential not treating the information provided by gesture as distinct from the information provided by speech.

Driskell and Radtke (2003) explored the relationship between gesture and comprehension in dialogue. In their study, one participant (the speaker) was tasked with conveying the meaning of a word but were not allowed to explicitly name it. The other participant (the listener) was required to guess the word. The study included two conditions, determined by whether or not participants were allowed to gesture. Driskell and Radtke (2003) used the number of questions a listener asked in response to each description as a measure of comprehension. The number of questions the listener asked was

used as an index of comprehension difficulty. Driskell and Radtke (2003) found that using gesture facilitated comprehension. Furthermore, although gesture made it easier for speakers to produce utterances, the effect of gesture on comprehension was still shown even when its effect on production was controlled. This final point demonstrates that the benefit of gesture is not tied a facilitatory effect of gesture on speech production.

The studies just outlined all suggest that gesture affects comprehension. If this is the case, then it is important to ask how it is integrated with speech. In two studies, Kelly and colleagues (Kelly, Özyürek, and Maris, 2010; Kelly et al., 2015) explore the integrated-systems hypothesis, which states that gesture and speech mutually and obligatorily interact with each other. In two experiments, Kelly, Özyürek, and Maris (2010) used a priming paradigm to investigate whether or not incongruent information presented in either speech or gesture affected comprehension. For example, if the prime showed a video of someone chopping vegetables, the associated baseline targets would be the word “chop” and a gesture showing someone miming the act of chopping vegetables with a knife. Incongruency is on two levels: weakly incongruent, e.g., “cut”; and strongly incongruent, e.g., “twist”, and there would be two levels for gesture: weakly incongruent, e.g., a gesture in which the middle finger and index finger represent the blades of a pair of scissors; and strongly incongruent, e.g., a gesture in which the hands enact the motion of twisting off the lid of a jar. Participants were asked whether or not either the speech or the gesture matched the prime. Kelly, Özyürek, and Maris (2010) found that participants were slower to respond and produced more errors when incongruent information was presented (regardless of modality) compared to the baseline. Further, they found that there was an effect of congruency strength on accuracy regardless of modality. However, they did find that participants were quicker to respond to speech targets than gesture targets. This suggests that speech may potentially be processed faster than gesture. In a second experiment, Kelly, Özyürek, and Maris (2010) repeated their first experiment except this time they only asked participants whether speech conveyed information relating to the primes. Gestures were still either baseline (congruent), weakly incongruent, or strongly incongruent. Kelly, Özyürek, and Maris (2010) show that incongruency in gesture has a negative effect in terms of accuracy, however it does not

affect response times. These two experiments, therefore, demonstrate that gesture has a mutual and obligatory effect on the processing of speech. Kelly, Özyürek, and Maris (2010, p. 266) use these findings as evidence for the integrated-systems hypothesis, suggesting that while most previous studies have treated gesture as a context for speech, they “have shown here that speech is also a context for gesture, suggesting that the two modalities co-determine meaning during language comprehension”.

Kelly et al. (2015), built on Kelly, Özyürek, and Maris (2010) by exploring whether or not the incongruency effect shown for gesture can be observed with actions. For this study, lexical primes (e.g., “scrubbed”) were followed by congruent audio “scrubbed the dishes” or incongruent audio “chopped the vegetables”. Accompanying the audio were congruent actions (e.g., video showing someone scrubbing dishes) or incongruent actions (e.g., video showing someone chopping vegetables) in one condition, or a congruent gesture (e.g., video showing someone miming washing dishes) or an incongruent gesture (e.g., video showing someone miming chopping vegetables) in the other condition. As with the previous study, Kelly et al. (2015) found that incongruent gestures affected the speed and accuracy of whether participants thought a target was related to the prime. They also found that action depictions had a similar effect. However, there was an additional finding that actions were identified more accurately, but gesture targets were more disruptive to the processing of speech. Kelly et al. (2015, p. 522, emphasis in original) explain these findings by suggesting that “even though gestures are visually less informative than action, they may be treated as *communicatively* more informative in relation to the accompanying speech. In other words, although gestures are stripped of much of the visual richness of actions, something important remains”.

If words prime gestures, then it might also be the case that gestures prime words. Yap et al. (2011) explored whether or not iconic gestures (not accompanied by speech) primed related lexical targets in a lexical decision task. In experiment 1, they found that gestures primed related words, suggesting that there is a close relationship between the processing of gesture and lexical items. However, because the length of time the gesture video was displayed, which led to a sizeable difference between the offset of the prime and the onset of the target, Yap et al. (2011) argue that they cannot rule out the fact that

participants label each gesture prime in their mind, leading to the priming effect. To combat this, in experiment 2 Yap et al. (2011) minimise the length of the gesture to limit to the potential for such implicit labelling. In this second experiment, gestures still primed related lexical items, however, the priming effect found in experiment 1 was stronger. This finding either points to the possibility that the gestures were not fully processed in the second experiment or that participants do, in fact, implicitly label the gestures and the priming is due to this labelling. So et al. (2013), building on Yap et al. (2011), explored the effect of gesture when it accompanies speech, since this is more naturalistic. They compared the effect of gesture on its own, speech on its own, and speech accompanying gesture on reaction times to related or unrelated words. While there was a priming effect in all conditions, So et al. (2013) did not find significant differences in terms of reaction times, but they did find a significant difference in terms of priming effect. The gesture-accompanying-speech condition and the speech-only condition demonstrated similar results in terms of priming effect. However, the priming effect found in the gesture-only condition was significantly greater. So et al. (2013) interpret these results as suggesting that the facilitatory affect of co-speech gesture occurs at a higher level of processing, involved in understanding the conceptual meaning of words, rather than at a lower level, involved in discriminating words from non-words. This argument would also account for the reduced priming effect found in the Yap et al. (2011) study, since the conceptual processing associated with this higher level might be associated with longer temporal latencies.

It is widely recognised that utterance producers gesture to varying degrees. Therefore, it might be the case that comprehenders have varying abilities to comprehend gesture. In a series of priming experiments, Wu and Coulson (2014) demonstrated that a greater visuo-spatial working memory capacity (either natural or induced) correlated with the comprehension benefits of gesture. This was not the case for verbal working memory. Wu and Coulson (2014, p. 49) suggest that this finding highlights the possibility that gestures “promote image-based simulations of the meaning of an utterance”. These image based simulations might therefore explain the benefits to comprehension afforded by co-speech gesture.



Gesture has also been shown to affect recall. Several studies have explored the effect of gesture on the recall spoken and spoken and gestured events. Cohen and Otterbein (1992) presented participants with unconnected sentences presented either with pantomime<sup>8</sup>, non-pantomime, or no gesture followed by an open-ended recall task. They found that both gesture types facilitated recall, with a greater proportion of detail being recalled. They also found that this effect was greater for pantomime than non-pantomime gestures. However, they did not find this effect when participants were presented with narratives instead of unconnected sentences. In a similar study, Feyereisen (2006) explored the difference between representational and non-representational (e.g., beat gestures) on recall. Feyereisen (2006) found that representational gestures aided recall more than non-representational ones. Moreover, in a second experiment they added a gesture to a sentence that was previously produced without a gesture. They found that the addition of representational gestures, but not nonrepresentational ones, suggesting that it is the meaningfulness of gesture that is important for recall. Galati and Samuel (2011) explored the relationship between gesture (congruent, incongruent, and none) and delay from presentation (short, intermediate, and long) on the recall of information presented through narratives. They found that target information was recalled more accurately in all three delay conditions when congruent gestures are presented alongside than incongruent or when gesture is not presented. Galati and Samuel (2011, p. 421) suggest that the presence of congruent gesture “protected against the effect of delay”. Therefore, the results of Galati and Samuel (2011) suggest that as well as affecting recall of unconnected sentences, gesture affects the recall of information found in narratives. This is true for both pantomime and representational gestures.

In a meta-analysis of experimental studies, Hostetter (2011) asked a series of questions specifically designed to show whether there was a general tendency for gestures to perform a communicative function. Hostetter (2011) found a series of illuminating results.

- Listeners have better comprehension of speech when it is accompanied by gesture.
- Gestures that accompany abstract topics do not significantly benefit communica-

---

<sup>8</sup>pantomime is what was labelled as mimetic enactment according to Enfield’s typology.

tion, however those that accompany spatial or motor description do have a significant benefit.

- Gestures are more communicative when they convey non-redundant information.
- The benefit to comprehension is not a by-product of the benefit gesture affords the speaker.
- There is no significant difference between scripted and spontaneous gestures; they affect comprehension in an analogous fashion.

Taken together, these findings (and the results of the other studies reported in this section) strongly suggest that gesture aids the comprehension of utterances. However, Hostetter (2011) excluded neuroscientific studies from her meta-analysis. Because part of this thesis concerns the real-time comprehension of gesture, it is worth reviewing the neuropsychological evidence for gesture's affect on comprehension.

### Neuropsychological Studies

Several event-related potential (ERP) studies have explored the effect of co-speech gesture. The methodology which is designed to measure ERPs is electroencephalography (EEG) (Hinojosa, Martín-Loeches, and Rubia, 2001). ERPs represent electrical responses of the brain, time-locked to a particular observation. These responses are typically characterised as being either positive or negative, reflecting their polarity, and according latency (Hinojosa, Martín-Loeches, and Rubia, 2001). In relation to language processing it has been robustly observed that a particular negative wave occurs at around 300-600 milliseconds during the comprehension of semantic information that is incongruous for the current syntactic context (Kutas and Hillyard, 1980). In their experiment Kutas and Hillyard (1980) presented participants with sentences that either ended expectedly or unexpectedly. Taking, for example, the sentences:(i) “it was his first day at **work**” and (ii) “he spread the warm bread with the **socks**” it should be clear that while “work” is predictable within the syntactic context, “socks” is not. When participants were presented unexpected sentence final words it was observed that this negative wave emerges at about 400 ms following the onset of the last word. This negative wave has become

known as the N400 and is particularly useful for investigating the semantic nature of communicative elements. Further, because of its relation to semantic processing the ERPs can be used to measure many hypotheses surrounding gesture's contribution to linguistic processing.

Kelly, Kravitz, and Hopkins (2004) presented one of the first experiments to demonstrate that gesture affects ERPs to speech. To do this they presented participants with audiovisual stimuli for which the gestural component either matched, mismatched, or complemented the semantic information of speech when referring to specific objects. Participants were then presented with two objects and had to select the object to which the actor within the stimulus video was referring. They found a large N400 effect in trials where speech and gesture mismatched, and that matched and complementary gestures produced different ERPs suggesting gestures do perform a semantic function during reference resolution (Kelly, Kravitz, and Hopkins, 2004, p.258). Wu and Coulson (2005) conducted a similar study, except they presented participants with cartoons followed by video recordings of individuals performing gestures whilst describing those cartoons (with the audio removed). The gesture presented information that either matched or mismatched with an element of the cartoon. In this study it was demonstrated that gestures result in a strong N450 response, however Wu and Coulson (2005, p. 660) suggest that this is analogous to the N400 response.

Whereas the two studies described above presented participants with either single multimodal expressions or images and gestures presented in sequence, Özyürek et al. (2007) investigated the effect of gesture that either does or does not fit the syntactic context. To do this they conducted an experiment similar to the one described in Kutas and Hillyard (1980), with additional trials including gesture match/mismatch. For example, they presented participants with the sentence “he slips on the roof and rolls/walks down” with either an accompanying gesture depicting a rolling (matching) / walking (mismatching) movement. Therefore, there are four conditions depending on whether the gesture and/or speech match or mismatch the first conjunct. In doing so, Özyürek et al. (2007) investigate what they refer to as a local match/mismatch (i.e. the gesture is (in-)congruous with the co-articulated lexical item) or a global mismatch (i.e. the

gesture is (in-/)congruous with the syntactic context), thus providing much broader evidence regarding the effect gesture has on comprehension. These two mismatches are derived from local and global integration of speech and gesture. Özyürek et al. (2007, p. 607) state that their aim is to “reveal the underlying nature and time course of these two types of multimodal integration processes”. This experiment is of critical importance because it affords the analysis of single semantic elements in speech and/or gesture. Özyürek et al. (2007) found that when speech and gesture mismatch the syntactic context there is the N400 effect associated with semantic incongruity. Thus both speech and gesture are immediately integrated into context. Furthermore, there was no difference between whether speech and gesture were mismatched individually (global and local mismatch) or whether they were jointly mismatched with context (global mismatch and local match). They take this to suggest that during comprehension language and gesture are not first combined into a single semantic unit before being assessed within a preceding context, but rather processed in parallel Özyürek (2002, p. 613).

In three experiments, Holle and Gunter (2007) investigated the effect of gesture on the comprehension homonyms, which were unbalanced in terms of frequency (e.g., “Ball” used to play sport > “Ball” at which one might dance). The homonyms appeared alongside gesture, which depicted one of the readings. High frequency words are referred to as dominant and low frequency words as subordinate. Initial clauses, which included the homonym, were followed by a second clause that disambiguated it. In experiment one, participants were asked to judge whether gesture and speech had been compatible. In experiment two, stimuli were followed by silent video clips of gestures (taken from experimental stimuli) or a visually presented probe word. Participants were required to judge whether or not probe gesture or probe word had been present in the experimental stimuli. Lastly, in experiment three they added another category of stimuli that contained gestures referred to as manipulators (cf. Ekman and Friesen, 1969). Manipulators are typically non-communicative gestures that involve the producer manipulating something in their environment (e.g., playing with their hair) and thus should not have the same effect on comprehension as semantically rich iconic gestures. The experiment was a straight forward replication of experiment two with the addition of

manipulators. In all three experiments Holle and Gunter (2007) observed that gestures compatible with the dominant reading of the homonym resulted in N400 effects when incompatible with sentential context, thus suggesting that gesture may play a role in disambiguating homonyms.

The overarching conclusion to take from these studies is that gesture and speech are tightly interconnected and processed together. Furthermore, Özyürek et al. (2007) suggest that speech and gesture are not first merged and then comprehended in a moment by moment fashion, but processed in parallel. This seems to be in line with McNeill (2012), and his argument of co-expressivity. Therefore, it seems that early gesture scholars were accurate and that gestures are comprehended as part of the utterance. However, a pragmatic perspective on gesture not only needs to explain whether or not something is comprehended, but also whether or not it is produced intentionally. The next section focusses on the production of gesture.

### **2.2.5 Gesture Production**

In order to fully understand the role of gesture during communication it is important to have a model of how gestures are tied to speech in the process of speaking. A rich area of research in this domain has been created by building on Levelt's (1993) speaking model. Levelt's model, depicted in figure 2.3, provides a serial processing model of speech production.

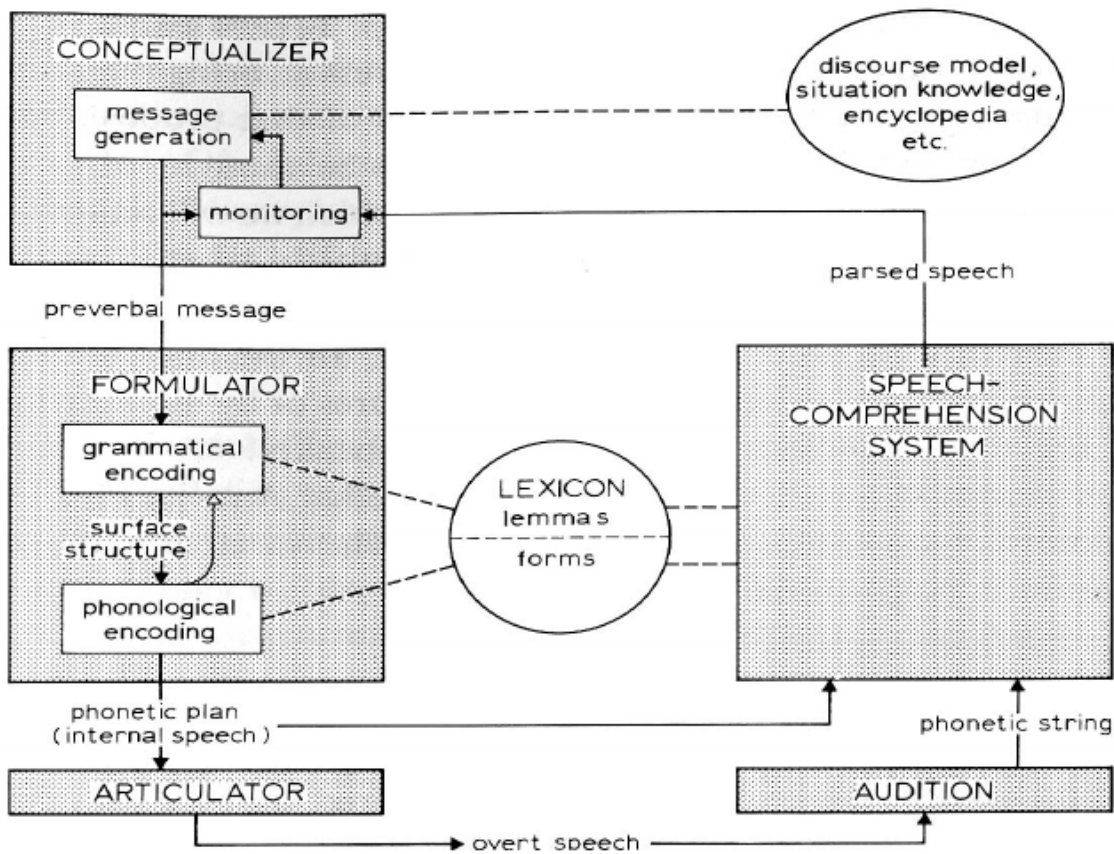


Figure 2.3: Levelt's (1993, p. 9) Speaking Model

According to Levelt's model, in the early stages there is a conceptualiser and a formulator. The conceptualiser is responsible for outputting the pre-verbal message (which is presumably in propositional form). The conceptualiser has access to broader aspects of cognition, including the discourse model, situational knowledge, and encyclopaedic information. The next stage, the formulator, takes the preverbal output of the conceptualiser and outputs linguistic form. The formulator has access to the speech comprehension system. The output of the formulator is not the token linguistic behaviour, but an internal (mentalist) representation of it. This internal speech goes back into the speech comprehension system both directly and once it has been produced. This means that the linguistic output is monitored at two stages, pre- and post-articulation. The conceptualiser then has access to the parsed speech of a speaker's own output.

Within the gesture literature there has been an attempt to represent the process of producing gesture alongside speech. There have been three main theories formulated. It is worth spending time to explore each and describe how they model the role of gesture.

Starting with Krauss's lexical access model (Krauss, Chen, and Gottesmann, 2000),

shown in figure 2.4a. The lexical access model built on the idea that the primary function of gesture is not to communicate. In other words, meaning in gesture is not part of a communicator's communicative intention. In the lexical access model, the pre-conceptualiser stage has been broken in two. First, long term memory is the same as in Levelt's model. However, working memory is now included and subdivided into propositional and spatial/dynamic working memory. Propositional working memory feeds into the conceptualiser and spatial/dynamic working memory feeds into a spatial/dynamic feature selector. This selector represents elements of objects abstractly and ultimately generates gesture. The important thing to note is that the process of generating gestures is triggered by the auditory monitor. Therefore, gestures emerge as a response to a producer's own monitoring of their speech and, as such, do not have access to grammatical form or preverbal messages. Once the lexical item that is related to the gesture has been produced this sends a message to stop the gesture. This model is called the lexical access model, because when a producer is struggling to produce a particular lexical item, features of a gesture may help facilitate that item.

Next, de Ruiter's (de Ruiter, 2000) model has been referred to as the sketch model but has been discussed in relation to the tradeoff hypothesis (de Ruiter, Bangerter, and Dings, 2012a). In this model, shown in figure 2.4b, the conceptualiser, which has access to both spatial and propositional working memory, produces two outputs: a message and a sketch. The sketch becomes the gesture component and the message becomes the speech component of an utterance. What's more, the gesture planner signals to the message generator, meaning that message generation is partially determined by planned gesture. Here, speech and gesture evolve from the same communicative intention, but represent different features of it. De Ruiter (2007) has referred to the sketch model as having a "postcard architecture" because, like a postcard, the sent message contains language and imagery. What's more, like a postcard, both elements are part of the message communicated. This model has been developed into the tradeoff hypothesis, which states that as speech become more difficult, people gesture more, and as gesture becomes more difficult, people speak more.

The final model, proposed by Kita (2000), has been referred to as the information

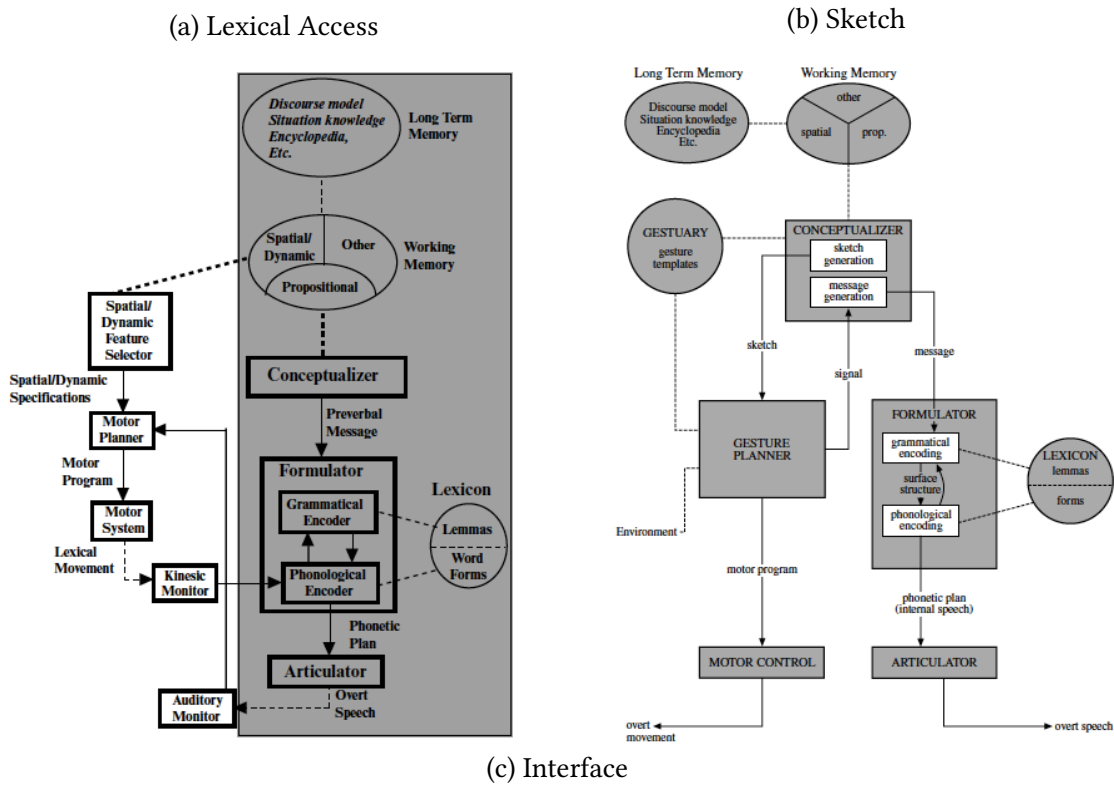


Figure 2.4: Three models of gesture production



packaging hypothesis and, more recently, the interface hypothesis (Kita and Özyürek, 2003; Kita et al., 2007). In this model, Kita and Özyürek (2003) distinguish from the two previous models, which they group under the term “free imagery” hypotheses. The interface hypothesis (see figure 2.4c) assumes that gesture has its origin in action and is represented as spatial-motoric information. Further, the interface hypothesis stipulates that there is a bi-directional relationship between message generation and gesture generation. To represent this, the conceptualiser has been divided in two, a communication planner and the lower, sub-divided action and message generators. The communication planner is responsible for communicative intentions and is responsible for which information is produced as speech and gesture and these are filtered over action and message generation. It is in the latter stage that the synchronisation of speech and gesture is determined. Another key feature is that formulator feeds back into message generation, which in turn feeds into action generation. Therefore, this model assumes that gesture is generated in a parallel fashion with speech—they are inter-generated. One of the key differences between Kita’s model and de Ruiter’s is that for Kita, the key process is the packaging of thought for language (or thinking for speaking, in the terms of Slobin (1987)). Therefore, whereas many theories assume that either gesture is generated prior to speech or that gesture is generated autonomously from speech, Kita assumes that gesture is fitted to linguistic structure. This perspective, because it assumes a strong inter-relationship between language and gesture has been called the hand-in-hand (So, Kita, and Goldin-Meadow, 2009) because it suggests that gesture and speech should regularly convey the same content.

These theories lead to different predictions. For Krauss, gesture should regularly convey non-communicative information. For de Ruiter, since speech and gesture are autonomous, then gesture should regularly be used to convey complementary information. For Kita, there should be a tight bond between gesture and language both typographically and in the local deployment of utterances.

The purpose in this thesis is to assess whether or not gestures are produced for communicative purposes. The first question then, is if gesture are not produced for communicative, then what are they produced for?. There are various suggestions.

It is possible that one of the key functions of gesture is to aid the cognition of the speaker. For example, gestures might aid lexical retrieval (Krauss et al., 1995; Morsella and Krauss, 2004) by representing aspects of lexical concepts. It is possible that one way in which gestures do this is by activating spatial imagery (Wesp et al., 2001; Morsella and Krauss, 2004). Gesture has been shown to lighten the cognitive load while people talk about mathematics (Goldin-Meadow et al., 2001), or promote recall (Wagner, Nusbaum, and Goldin-Meadow, 2004). Interestingly, Wagner, Nusbaum, and Goldin-Meadow (2004) found that gestures better aided recall when speech and gestures matched (presented the same information) than when gesture presented information additional to the information contained in speech. What is more, the effect demonstrated by gesture on memory is relevant to both spatio-visual and verbal working memory (Wagner, Nusbaum, and Goldin-Meadow, 2004). Finally, Wagner, Nusbaum, and Goldin-Meadow (2004) found that the recall of participants who naturally remained still (i.e., did not gesture) were not affected by being forced to remain still. Not surprisingly then, gestures have been shown to aid learning (Cook, Mitchell, and Goldin-Meadow, 2008; Novack and Goldin-Meadow, 2015).

Gesture is tied to additional motoric effort. It takes effort to produce a gesture. The above studies have demonstrated that gesture provides additional benefits that are not necessarily tied to communication. Similarly, gesture seems to reduce the cognitive effort involved in speaking (and not simply in terms of lexical retrieval (cf. Krauss et al., 1995; Morsella and Krauss, 2004)). Gestures, by their very nature are spatial (even though they benefit speakers in non-spatial ways). Therefore, it is possible that gestures benefit the conceptualisations underlying spatial thinking (Alibali, Kita, and Young, 2000). It has been shown that increased conceptualisation load, either by describing more complex objects or performing additional tasks, is tied to increased gesture (Melinger and Kita, 2007; Hostetter and Alibali, 2005). Further, Melinger and Kita (2007) argue that it is the additional load on selective attention that increases the production of gesture. So, in other words, it is choosing what to speak about that is one of the factors in gesture production. This fact is further highlighted by a study which shows that competing conceptual representations increased the production of gesture (Kita and Davies, 2009).

There are further suggestions that gesture may also highlight perceptually present information for speakers. Alibali and Kita (2010) found that children who were asked to describe spatial tasks (e.g., Piagetian conservation tasks) were more likely to describe perceptually present objects (using utterances such as “this one is tall and this one is short”). However, participants who were prevented from gesturing were more likely to describe perceptually non-present objects, by describing previous states (e.g., “they were the same length before”), hypothetical states (“if you put these two together, then this will be longer than this”) or transformations (“you moved it over”, “you didn’t add any”). This, Alibali and Kita (2010) argue, suggests that gesturing is linked to the highlighting of perceptually present information for speaking. However, the gestures analysed in this study seemed to be mainly deictic and modelling gestures that included a deictic component (e.g., point with a flat hand at a flat thing). Therefore, while these gestures seem to focus on perceptually present objects, it is likely that the gestures found in studies demonstrating more gestures occur when participants described spatial objects from memory than when participants can see the object (Morsella and Krauss, 2004; Wesp et al., 2001) are of a somewhat different sort.

It has also been demonstrated that people with high visualisation skills and low verbal (phonemic not semantic) skills gesture more (Hostetter and Alibali, 2007). These findings, together with the other studies already mentioned suggest that gesture is inherently linked to the conceptualisation of space. Furthermore, the production of gesture is likely caused by the process of packaging information for speaking. However, this says very little about the communicative nature of gesture and just because gesture aids cognition in various ways, does not preclude it having a communicative function, or even a primarily communicative function. It has been argued that speaking has a cognitive benefit for the speaker (Chomsky, 2010), as does writing or drawing (Dennett, 1993). To use Dennett’s computer metaphor, these processes free up RAM. There are several ways to explore the communicative function of gesture, and it is to these that we now turn.

One of the most fruitful, but debated topics in gesture studies is whether or not mutual visibility has an effect on gesture production. Famously, it has been demonstrated

that people still gesture when they are on the telephone (Bavelas et al., 2008) and that blind people gesture (Iverson et al., 2000; Iverson and Goldin-Meadow, 2001). Bavelas and Healing (2013) review fourteen studies using mutual visibility as a condition in gesture studies. Such studies are built on the premise that if mutual visibility affects gesture production, such that people gesture more when they can be seen, then gesture is being produced communicatively. However, if visibility does not affect gesture production then gesture is not communicative. Strikingly, Bavelas and Healing (2013, p. 65) report that seven of the fourteen studies report that gesture is affected by mutual visibility and seven find that it is not. When investigating this finding further, the single feature that separates those that demonstrate an effect of visibility versus those that do not is the use of quasi-dialogues (using confederates). The use of quasi-dialogues is tied to change in gesture when participants were mutually visible versus when they were not. In other words, in experiments that employed natural interaction, there was no effect of visibility. However, this finding is based on absolute gesture rate measure in accordance with speech. Bavelas and Healing (2013) find that when natural interaction is involved there is a difference, but it is not in terms of absolute gesture rate. Those features that do change with mutual visibility include the rate of gestures with an interactive function, pointing gestures, verbal and deictic references to gesture, and redundancy (lower when mutually visible). In other words, the overall finding of the mutual visibility studies seems to be that gesture is communicative.

It has also been shown that during lexical resolution difficulties, gestures are more likely to occur in a face-to-face context than one without mutual visibility (Holler, Turner, and Varcianna, 2013). This suggests that even those gestures that are potentially aiding lexical retrieval are more likely to occur when they can be seen. Also, while not directly related to mutual visibility, Özyürek (2002) has shown that gesture is changed depending on where an interlocutor is and/or how many interlocutors there are.

Another avenue to explore the effect of gesture during speech production is to explore the effect on speech when gesture is prohibited. Graham and Heywood (1975) found that in a task where participants had to describe line drawings, prohibiting gesture significantly increased the amount of time spent pausing, using demonstratives, and

the amount of time spent talking about spatial relationships. Emmorey and Casey (2002) explored the use of gesture in a task where participants were required to describe where to place blocks with complex shapes, so that those blocks filled a puzzle. They found that when participants could gesture (and see their interlocutor) they were more likely to describe object orientation when solving the puzzle. This finding might suggest that spatial language is coupled with gesture. However, speech referring to direction (of rotation, for example) was less likely to occur in gesture when speech described direction than when it did not. However, Hostetter, Alibali, and Kita (2007) explored the prohibition of gesture when participants were asked to describe performing actions (e.g., tying shoelaces). They found that the prohibition of gesture did not reduce the number of spatial descriptions, however, when gesture was permitted participants used richer verbs (e.g., “put” vs “cross”). Furthermore, participants who were not allowed to gesture also produced more utterances beginning with “and”, which Hostetter, Alibali, and Kita (2007) argue is acting as pausing device. Therefore, this study suggests that gestured information does not complement spoken information, but potentially enriches the information conveyed through speech. Additionally, the production of gesture increases fluency. These findings are in line with Rauscher, Krauss, and Chen (1996), who found that spatial descriptions become less fluent when participants cannot gesture. Finally, in a study (Hoetjes, Krahmer, and Swerts, 2014) where participants were required to describe different ways of tying ties it was found that being able to gesture did not affect speech duration, speech rate, or filled pauses. Therefore, the evidence that gestures have an effect on speech production seems inconclusive. However, there does seem to be a distinction between whether or not people are describing actions (e.g., tying shoelaces) or objects (e.g., line drawings). Furthermore, all studies that explored semantic features, do seem to find a difference when gesture is permitted.

The next key measure of production is the distribution of information across modalities. Holler and Beattie (2002) found that certain types of positional information tended to be expressed through speech and other types through gesture. Holler and Beattie (2003b) demonstrated that gesture is used when speech is used to resolve lexical ambiguity. Further, it has been shown that gesture is used to convey size information when

that information is crucial for the message an utterance producer is trying to convey (Beattie and Shovelton, 2006). Several studies (some described below) have shown that gestured information often mimics information contained in speech. However, Cohen, Beattie, and Shovelton (2011) demonstrated that when speech and gestures are analysed at a semantic level, 81.8% of gesture contained at least one semantic feature that was not present in speech. It has also been shown that even when people were allowed to choose whether or not they gesture, they frequently provide important semantic information gesturally (Melinger and Levelt, 2004).

The above studies all seem to point to the fact that certain semantic features in gesture are more likely to be produced when there is a possibility for them to be communicative. Thus, this goes against Krauss's (2000) theories of lexical access. However, it is not clear whether or not these results favour the sketch model (and tradeoff hypothesis) or the interface hypothesis. In a series of studies designed to directly investigate the interface hypothesis, there is a large body of evidence demonstrating that gesture matches the syntactic structure of speech (Kita and Özyürek, 2003; Özyürek et al., 2005; Özyürek et al., 2007). For example, in languages where the manner and path information of an event are produced separately as opposed to conflatedly (e.g., separate: "the ball went down the hill rolling" vs conflated "the ball rolled down the hill"), people also produce separate gestures (Kita and Özyürek, 2003). This has also been replicated with English only speakers when they naturally describe path and manner separately (Kita et al., 2007).

Experimental studies focussing on the tradeoff hypothesis are rare. One key study (de Ruitter, Bangerter, and Dings, 2012b) explored the hypothesis that as speaking becomes harder, gesture is more prevalent. However, this study found little evidence to support this claim and instead suggested that their findings more strongly support the hand-in-hand (or interface) hypothesis.

Therefore, summarising these studies, it seems that gesture does have a significant effect on both the production and comprehension of utterances. Further, in terms of comprehension it seems that gestures are integrated at the earliest stage possible. In terms of production, it seems that the literature most strongly favours the interface hy-

pothesis, which suggests that speech and gesture are based on the interface between two types of thinking. These findings all point to the fact that gesture must be included in pragmatic models of communication because it is a crucial factor in both production and comprehension. In terms of production, however, the gesture theories described all seem to focus on the informational content of the utterance as the end product of the process of utterance production. However, as was suggested in section 2.1.2 above, the information conveyed by an utterance is not the final goal of the utterance producer. The final goal of the utterance producer is to provide evidence that a comprehender can use to work out why they produced the utterance they did. From this perspective, it is expected that both speech and gesture will not be perfect signals, packaging informational content. It is for this reason that the next section explores the few theories of gesture that have been proposed in the pragmatics literature.

### 2.2.6 Gesture from the perspective of pragmatics

As explained in the introduction to this chapter, every model of pragmatics since Austin (1962) and Grice (1989) includes a notion of *intention* and some principles that guide behaviour in certain directions. Furthermore, these behaviours exist within a context, which afford producer and comprehender the opportunity to enrich the ‘meaning’ of that behaviour. The question of gesture, from this perspective, is whether linguistic behaviours and gestural behaviour form a single composite behaviour or whether they are two distinct behaviours produced together. This question may seem trivial, however, it is exactly the same question that has been explored in the studies above, it was just phrased in terms of whether or not gestures are communicative.

The contrast between gesture theorists and pragmaticists can be highlighted in a quote from Kendon (2004, p. 15):

deliberate expressiveness is *manifest*, it is perceived *directly*, and requires no deductive process leading to an *inference* of an intention. The intentionality of an action is something that is directly perceived. That is, it is the quality of the action as intentional (not the specific intention, necessarily) that is directly perceived. In other words an action that is gestural has an

immediate appearance of gesturalness. This means that a movement having this appearance will be discriminated and recognized as such directly. A detailed specification of what forms and movement patterns are required for a gesture to be discriminated remains, however, a matter for further work.

For Kendon it seems that gesture is interpreted because it is gesture and in order to understand the communicative value of gestures it is necessary to investigate the form and movement patterns of gesture. Kendon (2004, p. 15) goes on to say:

Whether an action is deemed to be intended or not is something that is dependent entirely upon how that action appears to others. [...] Actions can be varied so that they have more of the properties that will lead them to be treated as intentionally expressive, or fewer of them. This fact in itself is evidence that the judgement of an action's intentionality is a matter of how it appears to others and not a matter of some mysterious process by which intention or intentions themselves may guide an action may be known.

Kendon's arguments here miss a key point about the intentions studied in pragmatics. And, as Wharton (2009) argued in response to Kendon:

The aim of a cognitive pragmatic framework such as relevance theory is very much to engage with these 'mysterious' processes and examine the role they play. Indeed, one of the main achievements of Grice's work was to begin the demystification of such processes.

Kendon's argument may be an accurate description of how gesture is comprehended (i.e., directly), and that is a view that is open to debate. However, the driving impetus of pragmatic theory is that comprehenders of communicative acts do not perceive the meaning of the act directly, and that this is the case even with so called literal meaning. In fact, many contemporary theories of pragmatics assume that language grew out of the human capacity for intention attribution (Grice, 1957; Grice, 1975). From this perspective, the most parsimonious account of gesture's role in communication is one that explores its place within an intentional-inferential framework. Within the pragmatics



literature there are arguably two perspectives on the integration of gesture. These can be referred to as the *composite signal approach* (Clark, 1996; Enfield, 2009b) and the *gesture as a natural sign* approach (Wharton, 2009).

### **Gesture as a natural sign**

According to the gesture as a natural sign theory (Wharton, 2009), gesture is a natural sign, similar to a bear print, that communicates because it performs another purpose. For example, the bear print is a natural sign that a bear is/was nearby. The bear print is also a natural sign of the direction the bear was travelling in and the bear's size (amongst other things). However, these signs are not produced for the purpose of communicating anything, the print is simply a by-product of what the bear was doing (i.e., moving). This view, therefore, needs to explain what are gestures for. Wharton (2009, p. 153) suggests that the natural function of gestures is help speakers speak. Although Wharton is not clear on the details of exactly how gestures do this, the previous section outlined several benefits of gestures for speakers. Furthermore, because gestures regularly perform this function by analogously representing the content of what an utterance producer is speaking about, comprehenders are able to extract representational information from gesture. Furthermore, an utterance producer might be aware that comprehenders can extract meaningful information from gestures. In this case, the producer may choose to show the gesture to their interlocutor, who can infer the meaning of the gesture (which they could have done even if it was not intentionally shown).

This notion of showing is derived from Grice (1957) and can be explained using a non-gestural example. Coughing is a natural sign that is typically (not conventionally) associated with illness. Coughing can be taken as a sign that someone is ill. In the right condition, such signs can be deliberately shown. For example, someone who wants another person to know they are ill might ensure that they hear how terrible their cough is, thereby contributing to the meaning of their utterance (“I don't feel well”). Furthermore, an individual who knows that coughing is typically associated with illness can use this association to convey the fact that they are ill regardless of whether they are or not. Therefore, natural behaviours can be simply natural signs, they can be deliberately

shown natural signs, or they can be intentionally produced and shown natural signs. These distinct behaviours may result in distinct forms of comprehension. It seems unjust to say “you’re faking” to someone who is coughing due to genuine illness, and it seems unjust to say “stop milking it” to someone who has not deliberately shown a cough. For Wharton, iconic gestures are natural signs. He states, a “better understanding of the role of gestures in non-verbal communication may be gained by making use of the idea that some ‘natural’ gestures (in particular, ‘gesticulations’) are deliberately shown, even if they have not been intentionally produced” (Wharton, 2009, p. 152). What are they natural signs of? Wharton (2009, p. 153, italics in original) argues that “gesticulations are better treated as natural *signs* of the speaker’s desire to help the speaker understand”. In other words, Wharton treats gesture as being a result of a speaker’s reflexive thought, helping them communicative. This view is similar to what De Ruiter (2007) calls a *window architecture* in which gestures provide comprehenders with access to a producer’s thought process.

Wharton’s view seems to provide gestures with a ‘back door’ into the intentional-inferential communication process. This view is entirely consistent with all of the findings from the gesture literature focussing on comprehension, but it has a hard time explaining why certain semantic features are more likely to be produced in face-to-face contexts. Since the showing of a gesture is an intentional act, it is possible that speech production is modified to fit the gesture. Therefore, it would be expected that speech production would be affected by gesture production.

One of the weaknesses of Wharton’s view is that it proposes two mechanisms. Gesture is produced as a natural behaviour, whereas speech is produced as an intentional behaviour. Therefore, if Wharton’s theory cannot explain facts that can be explained by theories that suggest a single mechanism, then those theories should be preferred. The next section outlines such a theory.

### **Gesture as a composite signal**

The composite signal view is often attributed to (Clark, 1996, Chapter 6), who argued, following Peirce (1955; 1998), that signs may come in three forms. Clark built on Peirce’s

model by describing ground in relation to utterance producers, suggesting that we can define a signal “as the presentation of a sign by one person to mean something for another” (Clark, 1996, p. 160). From this perspective, different signs relate to different methods of signalling (summarized in figure 2.5).

<b>Method of Signalling</b>	<b>Sign Created</b>
demonstrating a thing	icon
indicating a thing	index
describing as a type of thing	symbol

Figure 2.5: Clark’s (1996, p. 160) ways of signalling

In line with what we said above, “demonstrating a thing” and “indicating a thing” fall under the category of non-conventional signals, whereas “describing as a type of thing” is a conventional behaviour. While useful conceptually, such stark distinctions are not tenable since signals are typically not monolithic behaviours consisting of a singular ground but are composite behaviours involving various semiotic relationships (cf. Clark, 1996; Enfield, 2009b). For example, an individual may point at a book and say:

(2.6) I love him

The gestural component of this utterance represents an indexical relationship between the gestural sign (the point) and the referent (the book). However, the verbal expression also consists of different semiotic relationships. For example, the first person pronoun (“I”) and the referent (the speaker) are in an indexical relationship because what “I” indicates is contingent on who produced it. The verbal component of the expression relating to the first referent’s attitude (“love”) and the attitude it invokes are in a symbolic relationship because it is convention that dictates what the word “love” means. The last component (“him”) indexically relates the first two components with some other referent (perhaps the author of the book), hence the overall signal is a composite signal (i.e., indexical gesture + indexical and conventional speech). However, it is also possible to suggest that these different elements of the spoken expression without the additional gesture do not consist of singular semiotic relationships but are also composite signals.

Pronouns are actually symbolic indexicals (Levinson, 1983, p. 65) since the relationship between “I” or “him” is only indexical so long as the recipient has access to the convention that “I” refers to the speaker. Pointing gestures, like verbal deixis, are also symbolic. For example in the western context the extended index finger handshake is standard, whereas in certain East Asian communities (amongst others) pointing with the head and lips is commonplace (cf. Enfield, 2001). This mixture of symbolic and indexical elements is further demonstrated by the fact that most westerners will point behind them using their thumb (Calbris, 1990); whereas speakers of certain Australian languages will continue to use the index finger (cf. de Ruiter, 2000). In sum, almost all signals (whether gestural, verbal, or some combination) are composite signals made up of composite signs.

Enfield (2009; 2013) has developed this perspective, building a theory of meaning that focusses on the composition of composite signals. Enfield’s view, however, focusses almost exclusively on comprehension, which he refers to as *sign filtration*. Enfield’s view of sign filtration is built on his view of comprehension. Sign filtration is operant during the search stage and is built on several heuristics. Enfield (2009b, p. 16–17) proposes two heuristics that are crucial for processing speech and gesture. First, the *contextual association heuristic* states that signs that are associated in terms of proximity and temporality, and should be taken as being part of a “single signifying action” (Enfield, 2009b, p. 16). The second heuristic is the *unified utterance-meaning heuristic*, which states that “contextually associated signs point to a unified, single, addressed utterance-meaning” (Enfield, 2009b, p. 17). These two heuristics may seem to be identical, however, there is a nuanced difference between them. The contextual association heuristic states that contextually associated signs should be taken together. For example, if someone says “I don’t feel well” whilst simultaneously coughing, the coughing should be taken as being contextually associated with the spoken utterance. However, the coughing is not part of the utterance-meaning. The kinds of iconic gestures that form the focus of this thesis, are, from Enfield’s perspective, distinct from coughing in that they *are* part of utterance-meaning. Therefore, a comprehender applying these heuristics would be guided towards interpreting the meaning of speech and gesture as representing a single, underlying,

speaker meaning. In terms of production, one must explain why a producer decided to use the particular composite utterance and not a different one that may also satisfy a producer's informative intention. It has been suggested that an ideally designed composite signal may include different semiotic elements opportunistically (cf. Bangerter, 2004; Clark, 1996). This is the position of the tradeoff hypothesis. Clark also suggests three guiding principles for why a producer produces the particular composite they did and not some other one. These are: 1. the purpose<sup>9</sup> of the utterance producer 2. the availability of different signals and 3. the effort required to produce a particular composite. Following these principles, one would expect that the choice of composite signal that an utterance producer produces will minimise effort and be maximally geared towards satisfying a producer's purpose.

To summarise these two perspectives, both theories suggest that gestures will affect comprehension. However, one might expect that since, following Wharton's view, gestures are derived from a separate mechanism to speech, that they may be comprehended differently. In terms of production, Wharton's (2009) view, although distinct, seems to be most in-line with Krauss's lexical access model. However, the composite utterance view seems to be in line with de Ruiter's sketch model and tradeoff hypothesis. Neither theory seems to predict the assumptions of the interface hypothesis, since they both treat speech and gesture as distinct. Therefore, one of the key questions that need to be addressed from a pragmatic perspective is how to capture the findings of the interface hypothesis. A proposal to this effect will be presented in chapter 7 below.

## 2.3 Action is Laminated

Above, it was argued that communicative behaviours are the result of goal-directed decision making processes. The aim of these processes is to produce signals that reliably mean (for the comprehender) what the producer intended them to mean. In this sec-

---

<sup>9</sup>When developing a theory of intentional behaviour it is crucial to assumption that purpose and intention are synonymous. A purpose can be realised without an intention. For example a screwdriver or an elephant's trunk can be said to have/serve a purpose, but neither screwdrivers or elephant's trunks can be said to have intentions. Further, we can ask questions such as "what's the purpose of that" but not "what's the intention of that". In short intentions require cognition and purposes do not. Similar arguments have been put forward in the literature (see Austin (1966) Kockelman (2005)).

tion, the productive nature of communicative behaviours will be further explored, not in terms of referential meaning but in terms of how they contribute to the activities in which we find them. In line with this, we can think of the notion of composite signal introduced earlier, as not simply pertaining to those elements integral to the referential meaning of signals but also to the ways in which communicative behaviour index, amongst other things: the activities interactants are involved in (e.g., buying a sandwich, giving directions, conducting a funeral); the stance of the two participants (e.g., shop owner and customer, husband and wife, friends); previous interactions (including the present one); and the current communicative behaviour’s position within the conversation (e.g., question-answer). These elements are best described by Goffman (1981), who calls them *laminations*. He (1981, p. 143) states:

One clearly finds, then, that coordinated task activity—not conversation—is what lots of words are part of. A presumed common interest in effectively pursuing the activity at hand, in accordance with some sort of overall plan for doing so, is the contextual matrix which renders many utterances, especially brief ones, meaningful. And these are not unimportant words; it takes a linguist to overlook them.

This view is also espoused by Clark (1996), according to whom communicative acts—acts in which one person means something for some other(s)—are not autonomous, unilateral acts of an utterer uttering and an addressee addressing. Instead, for Clark, such acts are bilateral, joint actions in which both interlocutors must perform their participatory acts in order to reach a joint construal of the intended meaning of an utterance. In this sense, the action of communicating is itself a collaborative joint activity, involving the actions of (at least) two people. These two actions form the two phases of a communicative act and are referred to as the presentation phase and acceptance phase (Clark, 1996, p. 227):

<b>Communicative Act</b>		
<b>Participating Actors</b>	Utterer	Addressee
<b>Participatory Acts</b>	<i>Presentation phase</i>	<i>Acceptance phase</i>

Table 2.1: Communicative Acts

The presentation of some communicative behaviour requires the acceptance of that behaviour in order for its intended meaning to be jointly construed by the participants. This definition of communicative acts shifts the notion of understanding from being the point at which an individual (e.g., a hearer) understands what is being communicated (referred to as the off-switch in section 2.1.2 above) to one in which understanding is the point at which a *mutual* agreement has been made regarding the speaker’s intention sufficient for current purposes (Clark, 1996, in particular, p. 222–224)—turning meaning into joint meaning (Carassa and Colombetti, 2009). Furthermore, the presentation and acceptance of a behaviour is part of a laminated action. For example if Bob were to say “hello” in response to Anne saying “hello”, this would not just inform Anne that Bob has understood her intention, but also that he has performed his role in the greeting. Bob’s “hello”, therefore, may be seen as finishing the greeting, leaving Anne to decide how she wishes to continue. In this sense, communicative acts can be thought of as existing on several levels within what Clark (1996, p. 222) calls an *action ladder*:

Level	Utterer A’s actions	Addressee B’s actions
4	A is <i>proposing</i> joint project <i>w</i> to B	B is <i>considering</i> A’s proposal of <i>w</i>
3	A is <i>signaling</i> that <i>p</i> for B	B is <i>recognizing</i> that <i>p</i> from A
2	A is <i>presenting</i> signal <i>s</i> to B	B is <i>identifying</i> signal <i>s</i> from A
1	A is <i>executing</i> behavior <i>t</i> for B	B is <i>attending</i> to behavior <i>t</i> from A

Table 2.2: Action ladder involved in language use

The action ladder is a useful way to conceptualise communicative behaviour because it highlights the fact that communicative acts are not simple acts of meaning and understanding. Communicative acts exist on, at least, four levels. At level one, they are acts of executing certain behaviours that must be attended to, for example making noises or moving one’s body. However, at level two, if these behaviours are to be understood communicatively, they must be presented as signals which may be identified according to their ground. At level three, these signals must be recognised as meaning something on this particular occasion. Lastly, at level four, these actions must be taken up as part

of a joint project. The different levels of the action ladder are upwardly causal, meaning that evidence that one level has been accomplished provides evidence that all lower levels have been completed (Clark, 1996, pp. 147–148). Therefore, what an individual needs to know is that the top most level of action has been completed because this provides downward evidence that everything below it on the ladder has also been completed. What's more, we can treat each level using Peirce's notion of the three-part semiotic process (introduced in section 2.1.2). For example, at every level of the action ladder, if we take A's act as being a sign, then it is possible to take B's act as a interpretant. The object of each level is the fact that the execution corresponds to the attention, the presentation corresponds to the identification, the signal corresponds to the recognition, and the proposal corresponds to the consideration.

For example, imagine that Bob scratches his face during a conversation with Anne. If Anne took his behaviour as existing only at level one then she might arrive at the conclusion that Bob's face is itchy and/or that he has dry skin. However, if Anne took Bob's behaviour as a presentation of a signal (level 2) then she would be warranted in believing that it means something. At level three, we can easily construct two scenarios for what Bob might be trying to signal: (a) It might be that Bob wants Anne to know that his face is itchy and he is ostensively scratching it to inform her of this fact; or (b) it might be that Anne still has some of the remnants of the chocolate éclair she has just eaten on her face and Bob is trying to draw her attention to it. In these two scenarios Anne has taken Bob's behaviour as a signal and derived two different intentions underlying it. It is crucial that as an analyst unless we have evidence of Anne's recognition, we cannot definitively know which signal she recognised. In the other words, we are unable to assign a definite object to Bob's sign. Further, these two potential signals propose different joint projects at level 4. In (a) Bob might want Anne to get him some face cream or just show some sympathy. However, in (b) Bob might want Anne to rub the food off her face. These levels are important because during interaction people may not appropriately attend to a behaviour at all stages of the process. For example, Anne might not see that Bob has rubbed his face, or she might not identify that it was a signal, or she might not recognise the correct meaning, which would ultimately lead to considering



the wrong project. Due to downward evidence, Anne's acceptance phase provides Bob with evidence of what level of the action ladder has been reached, if any. If she looks blankly or continues what she was previously doing then Bob might infer that she has not attended to his behaviour. If it is clear that she has attended to his behaviour but has not identified it as a signal, then Bob might draw the conclusion that Anne has not inferred that it was communicative. If Anne does something Bob was not expecting, then Bob might infer that she has not recognised the meaning of his signal and thus is considering the wrong joint project.

Clark refers to this process of presenting understanding during a presentation phase as *grounding*, by which he means added to the common ground of the participants sufficient for current purposes (see, *inter alia*, Clark and Brennan, 1991; Clark and Schaefer, 1989; Clark and Wilkes-Gibbs, 1986). In other words, the point at which communicators accept that their goals have been reached is not all-or-nothing—either they understand or they don't—rather it is flexible and depends on how important some goal is to the participants. Importance, too, can differ at each level of the ladder and for each participant.

A question which comes out of the action ladder approach is: what are joint projects? It is relatively easy to reconstruct someone's intention to communicate because, as human analysts, we have access to the same apparatus as the communicators. However, without additional information it can be difficult to grasp the larger purposes of a communicative act. These larger purposes have been described under various guises. Wittgenstein (1958) famously referred to them as language games, Bara (2011) has called them behavioural games, and elsewhere they have been referred to simply as games (Carletta, Isard, and Kowtko, 1996; Isard and Carletta, 1995; Kowtko, Isard, and Doherty, 1993). However, here we follow Levinson (1979) who refers to the highest order element of purposeful action as activity types. They may be defined as:

I take the notion of an activity type to refer to a fuzzy category whose focal members are goal-defined, socially constituted, bounded, events with *constraints* on participants, setting, and so on, but above all on kinds of allowable contributions. Paradigm examples would be teaching, a job interview, a jural interrogation, a football game, a task in a workshop, a dinner party and so on [...]. Elements of the structure of an activity include its subdivision into a number of sub-parts or

episodes as we may call them.

(Levinson, 1979, p.368–369, emphasis in original)

These nested elements are Clarkian joint projects (or nested games), which are contributions to the activity within which the communicators find themselves. In order to expand on these notions, imagine John, wanting something to eat, going into his local deli, and the following interaction ensues:

- (2.7) **John**           Hi  
**Cashier**           Hi there  
**John**                Can I get a sandwich?  
**Cashier**           What kind of sandwich would you like?  
**John**                That one, please  
                          ((Pointing to sandwich behind counter))  
**Cashier**           That'll be £2  
**John**                ((Hands over money))  
**Cashier**           There you go  
                          ((Hands over Sandwich))  
**John**                Thanks. Bye  
**Cashier**           Bye

This example does not convey the full complexity of such activities, but it serves to exemplify important elements of the laminated nature of activity. Here, John's activity is driven by his desire for food, which acts as the stimulus for his behaviour. However, in order to satiate his desire he must engage another individual. In doing so, John is treating the cashier as one of his many problem-solving resources (Enfield, 2009a, p. 71). Indeed, the cashier will also be treating John as one of his since his livelihood depends on his business. It is clear here that John and the cashier have different goals, John's is for food whereas the cashier's is for financial gain, however, in order to reach these goals they must coordinate during a single activity. Coordinated activities of this sort develop in paired sequences called *adjacency pairs* (Schegloff and Sacks, 1973) or

*projective pairs* (Bangerter and Clark, 2003). The canonical projective pair is the question/answer sequence which consists of a first pair part of a question and a second pair part of an answer. In this example, one first pair part is “can I get a sandwich?”, which begins the joint project and projects an appropriate second pair part, in this case the appropriate response might be handing over a sandwich. In the example, however, this does not happen, instead the first pair part is followed by another question: “What kind of sandwich would you like?”, which is itself another first pair part of a sub project. Sub-projects are nested inside of projects and are typically required in order for the higher level project to be completed. Importantly, this response does something further: it signals to John, not that the cashier has misunderstood him, which might be the result if he had said “pardon”, but that he cannot close the project without further information. As described above, the difference between a misunderstanding and a non-canonical response relates to the level at which the problem has occurred in the action ladder: misunderstandings occur at levels 1–3 whereas non-canonical responses typically relate to level 4 and thus index an inability to consider the project. Therefore a non-canonical response provides evidence that the cashier has understood John’s intention but cannot fully consider his proposal since he does not have enough information to do so. The joint projects continue until John’s goal has been reached and the activity comes to a close.

This interaction may be represented as such:

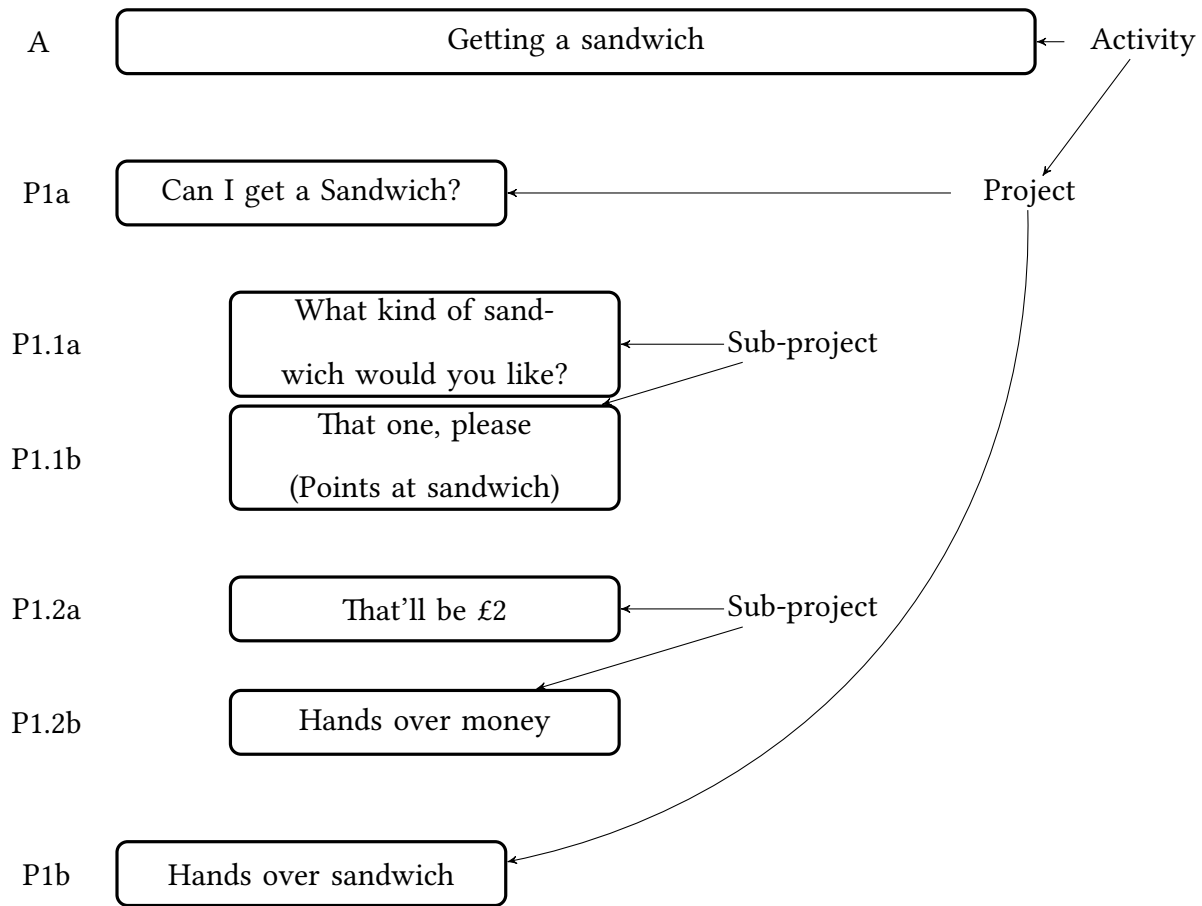


Figure 2.6: Activity Structure

Here, each project is represented using P followed by a number designating the level at which the project occurs during the activity. Also, “a” marks the initiating move and “b” marks closing move.

We can add one final Peircean element to this analysis in the form of chaining. One of the key insights of Peirce’s work is that a sign in one semiotic process can be an interpretant in another. Therefore, the cashier’s utterance of “What kind of sandwich would you like?” can be conceptualised as an interpretant of John’s utterance “Can I get a sandwich?”, which is a sign. However, it can also be conceptualised as a sign for which John’s utterance “That one, please” is the interpretant.

## 2.4 Summary & Research aims

In summary, I have suggested that it is better not to privilege the conventional (symbolic) nature of linguistic elements over other forms of behaviour when devising a theory of semantics and pragmatics, especially when the aim is to analyse gesture’s contribution to

meaning. In doing so, utterance production and comprehension are viewed as satisfying two distinct goals. The first goal relates to the understanding of the ‘content’ of an utterance, which can be inferred according to how well a behaviour relates to what an individual is taken as meaning within a specific context, what might be thought of as communicative competence. This correlates to level three on a Clarkian action ladder (see table 2.2). The second goal relates to how these communicative behaviours are fitted to and create the activity within which they are significant. This correlates with level four on a Clarkian action ladder. It was suggested that both of these can be approached using the Peircean notion of the third as an important conceptual tool. In other words, meaning at the different levels of the action ladder can be analysed as correspondance.

The central questions explored in this thesis pertain to how an individual is able to take a semiotically complex behaviour, such as an utterance consisting of speech and gesture, and understand it to refer to a single thing. Furthermore, why are the combinations which manifest themselves during interaction (for example, symbolic spoken words + iconic indexical gestural contributions) packaged in the way they are? Do gestures provide communicators with some advantage not typically afforded by linguistic only communication? Around these questions we can formulate two research questions:

1. What is the effect of gesture at level four of the Clarkian action ladder? This question gives rise to two sub questions:
  - (a) What is the effect of gesture on the proposal and consideration of joint projects?
2. What is the effect of gesture at level three of the action ladder? This too, raises two sub questions:
  - (a) For the producer, what is the effect of gesture on the composition of composite signals?
  - (b) For the comprehender, what is the effect of gesture on the recognition of composite signals?

The purpose of this thesis is to use the Clarkian action ladder to explore the effect of gesture at different communicative levels. The argument is that if gesture has an effect

on the distinct levels of the action ladder then this provides further evidence that gesture is part of communication proper. These questions will be addressed using different methodologies. Chapters 3 and 4 address question 1 and 2a. Chapters 5 and 6 address question 2b.

## **Part II**

# **Gesture, Meaning and Interaction**





# Chapter 3

## Methodology 1: The Map Task

### 3.1 Introduction

This chapter introduces the map task ((cf. Anderson et al., 1991; Anderson, 2006; Brown, 1995), section 3.2 below), which is first methodology used in this thesis. The map task is ideally suited to exploring the production and situated nature of composite utterances. This chapter and the next explore the production of composite utterances and how those utterances are operationalised during interaction. In terms of the Clarkian action ladder, the focus is split between *signalling*, *proposal*, and *consideration*. These different perspectives on composite utterances will be used to explore the two pragmatic perspectives on gesture described in chapter 2. According to the first theory (Wharton, 2003; Wharton, 2009), gestures are considered to be a natural sign. Therefore, gestures communicate incidentally, often because they serve another purpose. They can, like natural signals such as crying, be deliberately shown to an interlocutor and integrated into the production of an utterance. This suggests that gestures are not normally part of a speaker's communicative intention but may be co-opted by a speaker and may be communicatively shown. The alternative theory is that gestures are part of composite signals (not signs) and they are communicative for the same reason that other communicative behaviours are (e.g., spoken language) because their purpose is convey information to an interlocutor (Clark, 1996; Bara, 2010; Enfield, 2009a). From this perspective, speech and gesture are composed to be maximally communicative and are produced with the intention of guiding a comprehender to the informative intention of the producer. The reason why

one composite is chosen and not another is down to the producer's purpose, availability of behaviours, and effort in producing it. Therefore, from this perspective gestures and speech can be thought of two autonomous behaviours that have been selected to be put together. These distinct perspectives lead to several distinct hypotheses regarding use of gesture during interaction, these will be explored in more detail below.

However, one of the key elements that will be presented as part of the analysis below, is how semantic information is presented in both speech and gesture. Therefore, before continuing, it is worth exploring how the semantic content of utterances can be considered at two distinct levels. Taking each level in turn, first, there is semantic content that pertains to inherent properties of a referent, and second, there is semantic content that relates to a relationship between the referent and its place in the world. The section will begin with examples from two apposite studies.

Semantic explorations of variation in language has revealed fascinating insights into how languages (and potentially minds) categorise the world. Researchers have extensively explored diverse areas from the delineation of kinship systems to the labelling of elements in the colour spectrum. Here, there is not space to review this growing (although by no means new) field of research focussing on cross-cultural multimodality, but it is worth exploring some of the findings relating to a Paman language called Guugu Yimithirr.

Guugu Yimithirr has been the focus of Haviland (1993) for many years, however it was first described by the famous British explorer Lt. James Cook when, on the 11th June 1770, his ship, *The Endeavour*, ran aground on the Great Barrier Reef. It took seven days but Cook managed to secure his ship in the mouth of what is now known as The Endeavour River in Cooktown near Hope Vale and ended up staying there for almost seven weeks. In Cook's notes he describes how one of his seamen, after getting lost, was directed in the right direction by the speakers of Guugu Yimithirr. The fact that direction giving was recorded is significant because the way Guugu Yimithirr encodes space is the reason it has been included here.

Guugu Yimithirr is what Levinson (2003) would describe as a language with an Absolute spatial coordinate structure. Absolute systems employ a Global frame of reference,

using devices such as cardinal directions rather than using relative terms, such as “left” or “right”. Relative terms are underspecific and thus require enrichment in terms of perspective—it would usually be a mistake to assume that a speaker facing you means your right as opposed to their own. In order to understand a relative term it is important to understand the perspective taken by the producer. Absolute terms are not ambiguous in this way and a comprehender can understand an absolute spatial term without adopting an egocentric perspective. It is important to note that Haviland views the application of these categories (e.g., absolute and relative) as simplistic and not fully representative of Guugu Yimithirr’s complex spatial language (Haviland, 1993, p. 10), but it is, nonetheless, a useful tool to explore the language. Guugu Yimithirr’s spatial system categorises horizontal angles by employing one of four spatial roots: *gungga-*, *naga-*, *jiba*, and *guwa-* that correspond respectively to *north*, *east*, *south*, and *west*. A system using such terms would require a constant sense of geolocation, which, as will be demonstrated below, manifests itself multimodally. English, too, has words referring to geocentric cardinal direction, but they are rarely used in the way such words are used in Guugu Yimithirr where a speaker might utter:

(3.1) hv860718

naga=naga      mana-ayi

EAST+REDUP INCHO-IMP

*Shift a bit to the east!*

This utterance would be unlikely in most English contexts because the directional terms *right*, *left*, or a phrase making reference to another (secondary reference) object (e.g., *towards the road*) would typically be employed. However, the examples discussed below are interesting, not due to the presence of a cardinal direction term, but due to its absence. The example, referred to as “How the boat sank” appears in Haviland (1993), but the examples were recorded on two separate occasions. The first was recorded by Haviland in 1980 and the second was recorded by Levinson and Brown in 1982. Importantly, the examples do not include the use of Guugu Yimithirr’s directional roots. Both examples involve the same speaker speaking about the same event. The event de-

picts how the speaker swam away from a sinking boat, which capsized behind him as he headed for land. In the first video the speaker says:

- (3.2) 21j; dagu gulnguy nhayun . miidaarr-in yarrba gurra-y  
 thing boat that+ABS lift-PAST this=way say-PAST  
*Well, the boat was lifted up; it went like this.*

While producing this utterance the speaker simultaneously produces the gesture depicted in figure 3.1 which shows him bring both his hands up in front of him and then down in a rolling motion.



Figure 3.1: Taken from Haviland (1993, p. 15)

However, in the second recording the speaker utters:

- (3.3) 17j; miidaarr-in yarrba th—  
 lift-PAST this=way  
*It lifted it up like that—*

- 18 thambarr-in  
 throw-PAST

–and threw it

During the utterance this time the speaker produces the gesture depicted in figure 3.2, which shows him lifting his left arm while dropping his right arm and then producing circles inwards with both. The major difference between these two gestures is that in the first example, the direction in which the boat is rolling runs perpendicular to the trunk of the speaker's body, whereas in the second it runs parallel. Most people watching and listening to the speaker would have probably missed the directionality encoded in the gesture. However, if taken in coordination with the direction the speaker is facing while telling the story, it is possible to work out that both these gestures depict the boat rolling over from east to west. Interestingly, meteorological evidence suggests that winds at the time of the accident would have been blowing from the south-east (i.e., towards the north-west) and as such it is likely the actual direction the travelled in would be from east to west (Haviland, 1993, p. 18).



Figure 3.2: Take from Haviland (1993, p. 16)

This remarkable sequence provides evidence for the fact that speech and gesture are linked in terms of how the producer construes the world. The way speakers think about

extralinguistic features, such as frame of reference or whether a referent is in motion, have an effect on not only their linguistic encoding of events, but their gestural one as well. These features are tied to the way that utterances semantics is link to the ways in which interlocutors construe the relationship between an object and the world. However, speech and gesture also interact at level of inherent to the meaning of a particular referent. For example, different languages encode event description in different ways. A key difference explored in the literature relates to distribution of elements relating to trajectory and manner in motion events (Langacker, 2013; Talmy, 2000a). The following examples are taken from Kita and Özyürek (2003, p. 22) and show how English, Japanese, and Turkish represent certain motion events:

- (3.4) (1) He rolls down the hill  
 (2) Japanese  
 [korogat-te] [saka-o kударu]  
 roll-Connective slope-Accusative descend:Present  
 “(s/he) descends the slope, as (s/he) rolls.”  
 (3) Turkish  
 [yuvarlan-arak] [cadde-den iniyor]  
 roll-Connective street-Ablative descend: Present  
 “(s/he) descends on the street, as (s/he) rolls.”

In these examples there is a clear difference between the way these elements are encoded. In English, trajectory and path form a tighter linguistic unit than Japanese or Turkish. In an experimental study, Kita and Özyürek (2003) found that speakers of Japanese and Turkish speakers were more likely to accompany descriptions of motion events with gestures that only depicted manner or trajectory than English speakers. It has also been demonstrated that these processes are not just tied to different languages but are also relevant to different ways of describing an event within a single language (Kita et al., 2007).

These examples and the examples from Guugu Yimithirr highlight two different levels at which we can analyse the meaning of gesture and speech. The Kita and Özyürek (2003) study focusses on how the different features of the referent are presented through speech and gesture. The Haviland (1993) study instead focusses on the relationship between the referent and the producer's conceptualisation of the world. These two levels represent distinct features of how people represent the world through talk and gesture. The methodology presented in this chapter is designed to explore both.

The purpose of this study is to analyse the effect of gesture on the wider projects within which utterances are situated (Bangerter and Clark, 2003; Bangerter, Clark, and Katz, 2004; Clark, 1996). The map task is a method ideally suited to exploring this area since it was designed to scrutinise language use during a collaborative joint activity in which two individuals have to cope with diverging representations of the world.

## 3.2 The Map Task

The map task adopted in this thesis deviates from the classic map task (cf. Anderson et al., 1991; Anderson, 2006; Brown, 1995). However, before explaining how it differs, it is worth outlining the classic model. The map task (in particular those that make up the HCRC map task corpus (Anderson et al., 1991)) involves two participants—the information giver (IG) and the information follower (IF)—who both have a two-dimensional map. The Giver's map contains several landmarks, a start point, a finish point, and a route connecting the start and finish points. The follower's map has several landmarks and a start point, but does not contain a finish point nor a route. Additionally, there are several landmark mismatches, which may include landmark exclusions. The participants' task is to *physically* recreate (i.e., draw), as closely as possible, the director's route on the follower's map. Importantly, neither participant has direct visual access to the other participant's map. Therefore, in order to overcome the coordination problem created by the map task, the two participants must rely on a fractured and incomplete shared visual environment during communication.

It is because the follower keeps a physical record of the route they recreate that the researcher is able to ascertain a wealth of information regarding the follower's compre-

hension of the giver's instructions. Moreover, because the analyst controls the information presented on the maps, they have privileged access to what the giver is trying to describe. For example, if the participants have found the task particularly difficult, a researcher would be able to explore the follower's map and gain important information regarding if or when some confusion occurred. Furthermore, the fact that the situations in which the participants are involved are similar across trials places the analyst in a special position. As Brown (1995, p. 43) argues: "the strength of the analyst's position here lies in being able to find other examples of similar behaviour by other pairs of subjects confronting the same, or very similar, communicative problem".

Another interesting element of the map task is that the two maps have different landmarks: the giver's map will both be missing and have additional landmarks in comparison to the follower's. It is this carefully-manipulated element of the task that often produces diverging understandings of the maps, and it is this feature that must be overcome by the participants in order to complete the task. As noted in previous work (Wilson, 2011) individuals appear to rely on the visual modality as a stable communicative resource in both participants' immediate environment<sup>1</sup>. Put plainly, participants appear to superimpose their immediate visual environment into the gestural space. The result of this is that space abstractly stands in for the map, so that pointing in space is actually pointing at the map and a clenched fist held in front of the body may actually represent a landmark in abstract gesture space. All this makes the map task an environment, par excellence, in which to examine gesture.

The map task then is a suitable task for applying a semantic feature analysis, which according to Gerwing and Allison (2009a) seems to be the most useful method for understanding the contribution of both speech and gesture during reference in conversation. Below I will explain how speech and gesture were analysed using a semantic feature analysis.

---

<sup>1</sup>See Clark and Krych (2004) for a similar observation regarding a different task in which participants build lego models.



### 3.2.1 This Map Task

Due to the nature of the research questions addressed here, it is a requirement for the task to be more constrained than previous map tasks outlined in the previous section. First, the maps were controlled across all trials, so that all givers and all followers have identical maps. This meant that rather than focussing on either accuracy across entire tasks or accuracy for individual tasks, which are not generalisable, a focus could be placed on individual sub-projects within the experiments. As will be discussed below, these sub-projects are determined by the participants themselves, however, there does seem to be some consensus regarding their delineation during the experiment. Second, the landmarks were also controlled for syllable length so that any communicative difficulty related to an individual landmark is not attributable to the choice of landmark name.

Additionally, there were modifications that were designed to create the potential for particular phenomena typically observed during spoken interaction. These were included to test whether or not gestures are susceptible to local conventionalisation—a concept referred to as entrainment—which relies on two or more people forming a tacit agreement (or “conceptual pact”) regarding the manner in which they will refer to a specific item (Brennan and Clark, 1996; Brown-Schmidt, 2009). The focus of such studies is typically on lexical entrainment, however, here we are not interested in lexical entrainment but rather the conventionalisation of gesture form and the way in which collocation of a route shape and landmark affect the gesture produced. If a gesture’s form is repeated by *both* participants when describing the same object (cf. Yasui, 2013) then it provides evidence that that gesture has potentially been conventionalised. Further, this would provide evidence that (i) gestures were comprehended as part of a speaker’s utterance; and (ii) that something that was iconic, such as a tracing gesture, may become symbolic. These manipulations are not discussed in the analysis presented in the next chapter, but offer an avenue for additional research. These findings would be especially relevant with the collection of additional tasks.

### 3.2.2 Set Up

The physical set up of the room is presented in figure 3.3. Each map was placed on an easel designed to prevent each participant from seeing the other’s map. The easels additionally allowed for the attachment of a small camera (no. 3 in figure 3.3) to the top of the map, permitting close observation of the follower’s drawing. The remaining three cameras (1, 2, & 4) focussed on the giver, the follower, and the scene respectively. This allowed for the observation of participants individually and collectively. All recordings were made in a sound controlled room.

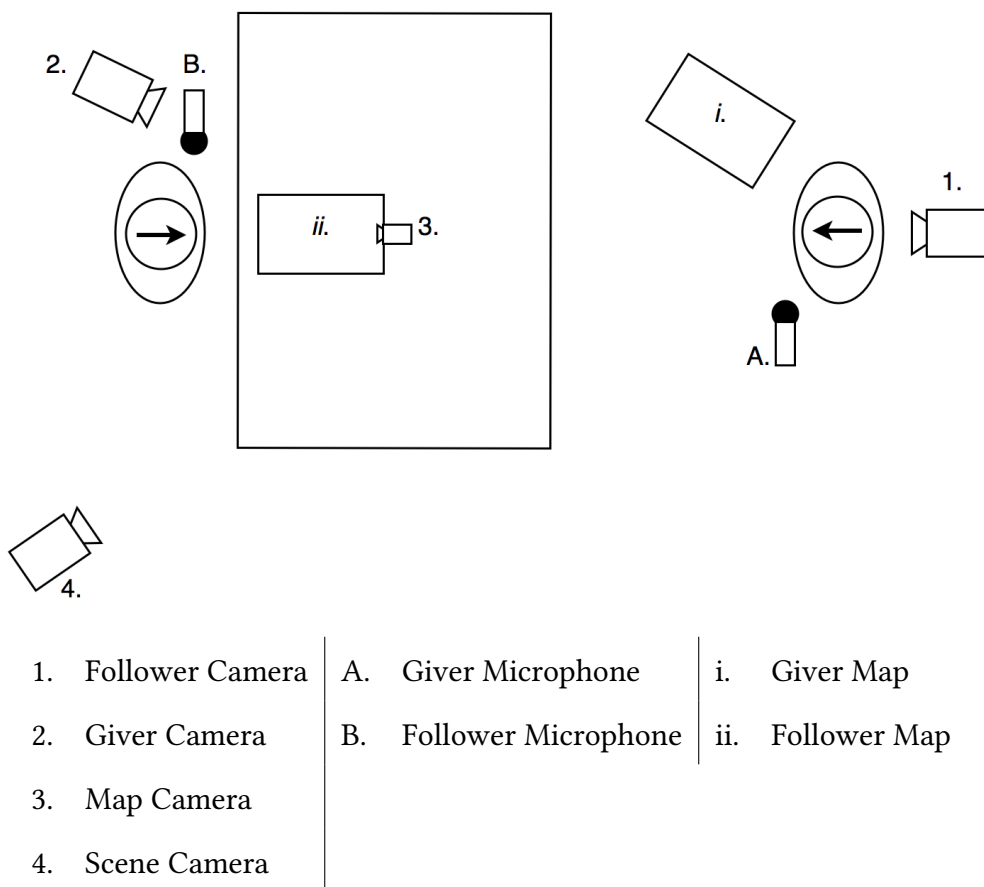


Figure 3.3: Physical Set Up

### 3.2.3 Maps

The maps (figures 3.4 & 3.5) have been designed with two interests in mind. The first is an understanding of the overall communicative success of the participants when describing the route shape. In this case the givers are presented with privileged information (i.e. route shape and finish point), which they must attempt to communicate to the follower.

Furthermore, in order to avoid a situation in which participants overuse the landmarks as anchors from which route shape can be inferred, landmark mismatches were included. The placement of mismatched landmarks is always the same (i.e., the placement of the giver’s pyramid = the placement of the follower’s old temple; the placement of the giver’s broken gate = the placement of the follower’s picket fence, see figure 3.5:A) so they still require some explanation on the part of the participants. The real purpose of these mismatches is to guide the participants into believing that their maps are less similar than they actually are thereby increasing the propensity for gesture.

The second important element of the maps’ design was the reduplication of landmarks and route shape, which appear in two environments: + route shape reduplication + landmark reduplication (see figure 3.4:A) & + route shape reduplication – landmark reduplication (see figure 3.4:B). The motivation behind this manipulation is to investigate whether participants produce similar gestures when describing identical route shapes and whether providing identical landmarks has an effect on this behaviour.

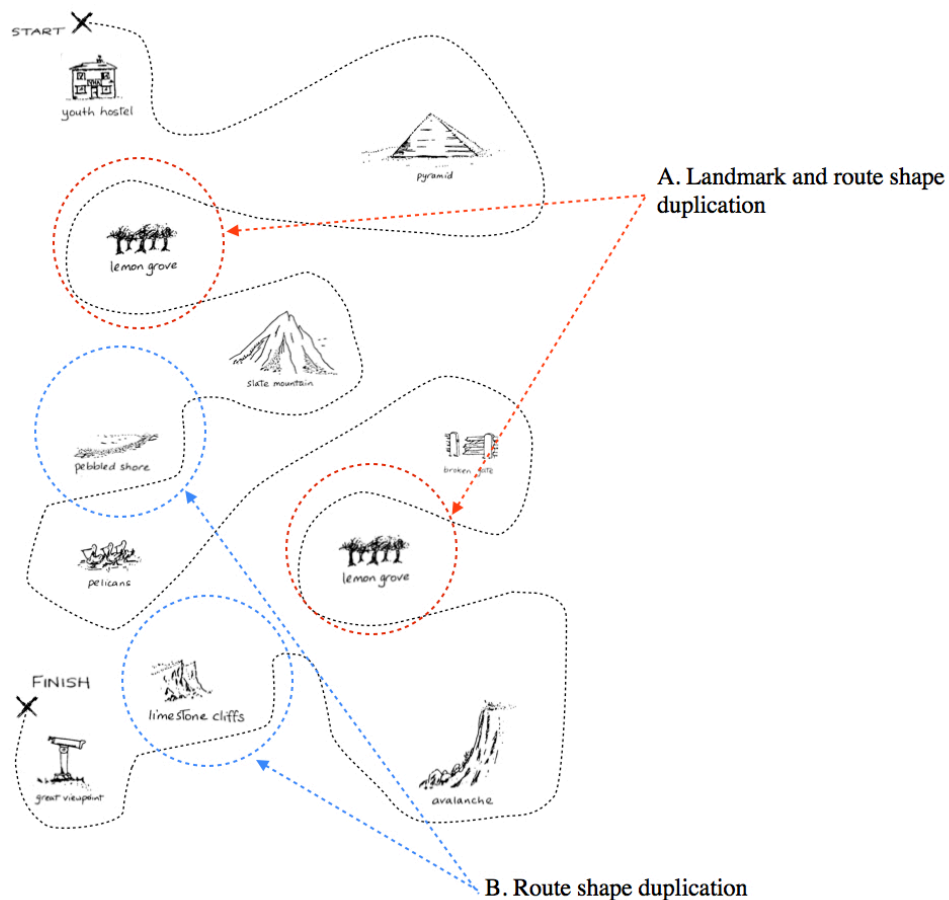


Figure 3.4: Giver’s Map

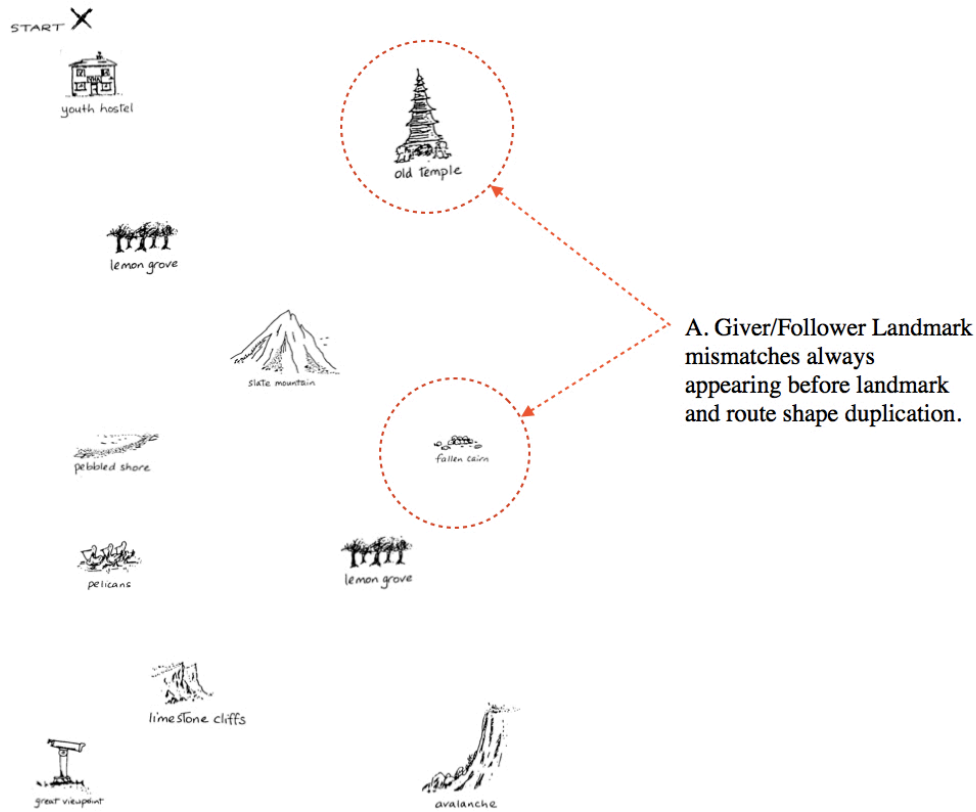


Figure 3.5: Follower’s Map

In summary, the task has been specifically designed to assess: (i) the overall communicative success of the participants during the task, from which the utility of their gestural contributions can be determined; and (ii) the extent to which both verbal and gestural elements may be conventionalised over the course of the task.

### 3.3 Coding

The data was coded according to projects and sub-projects, moves, and semantic content.

#### 3.3.1 Projects and Sub-projects

Projects were divided up using the same principles as those relating to the sandwich shop example (see 2). In this sense, projects and sub-projects are not something that the

analyst imposes on the data, but something that emerges discursively.

For the map task, the topmost project, referred to as the A-PROJECT, is the task itself. Within the A-PROJECT are B-PROJECTS that represent smaller goals such as “locating the lemon grove” or “going around the limestone cliffs”. Within B-PROJECTS are C-PROJECTS, typically formed by two moves (i.e., adjacency/projective pairs). A typical C-PROJECT includes an instruction to do something followed by an acknowledgment that the instruction has been understood. These first three stages are necessary, not just for the map task, but for most complex activities. However, D-PROJECTS (and so on) can be reached if a problem arises during a C-PROJECT. For example, if rather than following an instruction with an acknowledgement, which signals that it is appropriate to start the next C-PROJECT, a request for more information is provided, then this request would be an initiation of a D-PROJECT. These requests are nested inside the higher-level C-PROJECTS. This nesting of one project inside another is theoretically infinite, however, in the experiment described below no project goes below the K-LEVEL.

The analysis presented below only focusses on projects deeper than the C-LEVEL and the deployment of speech and gesture at the various levels. In order to ease analysis, a numerical system has been adopted. In this scheme C-LEVEL moves are coded as 1 with the move’s position within the project being labelled numerically. As such, the first move the first C-PROJECT would be labelled as 1.1 and the second move of the fourth C-PROJECT is labelled 4.2. This means that not only is it possible to analyse the vertical depth of a move, but also the horizontal position. However, because this study focusses on regularity across the map task, horizontal position is largely ignored because of the variability in the time it took participants to complete the task.

### 3.3.2 Move coding

Move coding is taken from *game coding* (Carletta, Isard, and Kowtko, 1996; Isard and Carletta, 1995; Kowtko, Isard, and Doherty, 1993). Game coding is an alternative to dividing tasks into projects and sub-projects, but it rests on the same theoretical premise—described above in section 2.3—namely, that interaction can be broken down into units and nested sub-units that break up the activity into manageable, discursively organised

chunks.

In game coding, utterances are referred to as moves, which either initiate a new game, respond to the initiation of a game, or act as preparation for the initiation of a game (Carletta et al., 1997). Furthermore, if we think of the map task as an activity type (Levinson, 1979), then, as with all activity types, it will come with a set of expected move types. Since game coding was specifically designed to analyse map task data it provides a useful way to categorise the participants' moves during interaction. These move types are broken into two main groups. The first group consist of those moves which initiate a project or sub-project:

1. **INSTRUCT**: Communicates a request or instruction for some action. Examples include “so you're gonna go down and under the pebbled shore”, “and then go round the pelicans”.
2. **CHECK**: Questions some previously given instruction. Examples include “am I circumventing something”, “so when you say inside you mean like the top of the trees”.
3. **QUERY-YN**, **QUERY-W**: Yes/No questions (**QUERY-YN**) or Wh-questions (**QUERY-W**) are questions pertaining to some unknown element. They are not requests for clarification since that would be a **CHECK**. They often form the preparation for future **INSTRUCT** moves. Examples include: “do you have the broken gate?”, “which way?”.
4. **EXPLAIN**: An explanation of the present situation or position. Importantly, it is freely offered and not elicited by a **QUERY**. Examples include: “so when you look at the line of your route it's almost like you missed out the pebbled shore”, “I'm under the pelican at the moment”.
5. **ALIGN**: Ensures interactants are aligned in terms of understanding or position. They often appear as tag questions or agreement tokens following long sequences of **INSTRUCT** headed (sub-)projects. Examples include: “that's the same as me”, “do you understand what I mean?”.

Of these, INSTRUCT and EXPLAIN do the majority of the work during the map task. CHECK, ALIGN, QUERY are responses that begin new embedded projects.

The second set of moves are those that typically serve the function of closing a minimal joint project. They often don't appear as full turns, but at the beginning or end of turns:

1. REPLY-Y, REPLY-N: Elicited response to a QUERY-YN, CHECK or ALIGN. They can be either positive or negative and sometimes include additional information. Examples include: “yeah, facing left”, “I don't have nothing that says pyramid”.
2. REPLY-W: Elicited response to QUERY-W, CHECK, or difficult to answer QUERY-YN moves. Example include: “it's parallel to the pebbled shore” and “you'll go right”.
3. ACKNOWLEDGE: The most common type of acceptance type move. Often referred to as back-channelling, they are acknowledgements not necessarily to take up the (sub-)project but that it was understood and that the presenter may continue. They do not have to be spoken, sometimes the follower's drawing is taken as an acknowledgement of an INSTRUCT move. Alternatively, ostensibly looking the map can be seen as an acknowledgement of understanding the content of a move but not being able to take up the project. Spoken examples include: “okay”, “right”.
4. READY: Indicates intention to begin new (sub-)project by focussing the attention on oneself. They are often placed at the beginning of INSTRUCT moves. Examples include: “okay”, “erm”, “yeah”. Eye-gaze can also serve this function.
5. CLARIFY: Follow CHECK headed sub-projects and generally reiterate a previous INSTRUCT move sometimes including additional information. Examples include: “round the outside of the page round the fallen cairn”, “towards the right”.

### 3.3.3 Semantic Coding

In this subsection, the semantic coding system used in this thesis will be outlined. This system forms the basis for the analysis of moves within the map task. The moves themselves are analysed for how they contribute to discourse, but the semantic coding is the

tool for understanding what they mean. Semantic units are analysed at a level beneath that of a single move (this will be explained in more detail below). Further, the semantic coding is divided into three sections: 1. The first describes the relationship between an utterance and the referent. This relationship is described in terms of semantic features; 2. the second describes the relationship between an utterance and the frame of reference adopted by the producer of the utterance; 3. the third describes the perspective of the producer as they produce the utterance. These three approaches relate to properties of the referent, properties of the conceptualisation of the world, and properties of the relationship between the producer and their body respectively. From this point on the coding of semantic features will be referred to simply as semantic coding.

### **Semantic feature coding**

Previously, the semantics of event representation have been explored in detail by Talmy (2000b), who attempts to break the semantics of space into an exhaustive typology. Using Talmy's typology and introducing some additional elements Loucks and Pederson (2010, p. 109) highlight the following ten categories:

1. MOTION: The fact of physical motion
2. PATH: The course followed by the figure with respect to the ground
3. FIGURE: The entity in motion
4. GROUND: The location with respect to which the figure moves
5. MANNER: The way in which the figure moves
6. CAUSE: The event that initiated the motion
7. ORIGIN: The origin of the path
8. ENDPOINT: The end of the path
9. RECIPIENT: the animate entity at the end of the path who receives the moving entity
10. AGENT: the animate entity which caused the motion

This typology can be understood through a sentence such as “John bounced the ball from one side of the room to the other where Sally was standing” which can be described as having the following semantic structure:



1. PATH: the ball is in motion
2. PATH: from one side of the room to the other (i.e. across the room)
3. FIGURE: the ball
4. GROUND: the room
5. MANNER: bouncing
6. CAUSE: John's bouncing of the ball
7. ORIGIN: One side of the room John
8. ENDPOINT: The other side of the room Sally
9. RECIPIENT: Sally
10. AGENT: John

This kind of typology is excellent for describing every motion event one can think of. However, this study focuses on route descriptions, which in addition to being a particular sort of motion event, often represent fictive rather than factive motion because the thing being described (i.e., the route) is not actually in motion (Talmy, 2000a; Streeck, 2009). A slightly different typology has been established for the semantic analysis of speech and gesture units. The ten semantic categories are described in Holler and Wilkin (2009) (see Gerwing and Allison (2009b) for a review of this methodology) have been developed to explore the presence or absence of semantic information in speech and/or gesture. These categories presented in Holler and Wilkin (2009, pp. 276f.) are:

1. *Entity (E)*: representation of the involvement/existence of an entity (speech examples: the kid, he, a car, it; gesture examples: a hand movement showing someone picking up an imaginary object from the ground; the hand representing the body of a car; a deictic gesture locating an entity in the gesture space).
2. *Identity (I)*: further information about an entity, such as whether the entity is animate or inanimate, as well as general category knowledge (e.g., that the entity is a human, or a type of animal; what category of inanimate entity, e.g., car, crops; speech examples: car, boy, man, owl, wheat, cane (but not: he, she, it, they, etc.); gesture examples: representation of specific features of an entity, such as wings, or particular clothing worn by a character in the story).

3. *Movement (MO)*: representation of movement, but not necessarily its (accurate) path nor type of action; speech examples: coming past (vs. driving past), to put (vs. to throw), to leave (vs. to run away); gesture examples: representation of general movement of an entity, such as in a lax, brief hand movement towards the right as if something is travelling out of the middle of the gesture space.
4. *Action (A)*: representation of the type of action taking place (e.g. grabbing, holding, walking, driving, poking, throwing, running, seeing, sitting). N.B. This was scored in addition to *Movement* when the type of action, by definition, also involved movement (speech) or this movement was gesturally represented. Actions such as ‘seeing’ were therefore not scored for *Movement* but for *Action* only.
5. *Direction/path (D)*: representation of the direction/trajectory/path of a movement or activity (e.g., up, down, straight, diagonal, a curvy line). This category also contained information about actions that were directed at other entities (D-O) (e.g., shouting *at* someone, saying something *to* the driver, shaking a stick *at* someone).
6. *Manner of action (MA)*: representation of the manner of action, i.e., the way in which an action is being carried out (such as, speech: walking *quickly*, waving the stick *angrily*, sitting *in a row*, coming out *one after the other*; gesture: representation of a speedy movement, representation of individual entities as being positioned in a line/row).
7. *Appearance (AP)*: representation of information about the appearance of entities (often adjectival, for example, looking *cute*, a *fluffy* owl), including information about their specific type (e.g., *barn* owl, *wheat* field).
8. *Position (P)*: representation of the position of one entity relative to another, or relative to its surrounding space (speech examples: the man got hit by the car; the kids hide behind the grass; gesture examples: representing a car with one hand, a man with another as well as the distance/contact between them). This semantic category also includes location information (P-Loc) (e.g., over there, in the distance, nearby).
9. *Size (S)*: representation of the size of entities (speech examples: a *little* bin,

*tall* grass, an *average-size* house; gesture examples: showing the width of a field, size of a window, body of an owl). This category also includes quantity information (S-Q) (e.g., some, little, all; quantity information expressed through singular and plural markers of nouns and pronouns [except uncountable nouns without a quantifier, such as ‘the litter’, which could be a small or a big amount, and the pronoun ‘it’, which could refer to both a single person/thing as well as a group of individuals/undefined quantity of an entity], and quantity information expressed in gesture [e.g. representing an entity as being a big group/more than one individual]).

10. *Shape (SH)*: representation of the shape of entities (e.g., outline of a square window frame, a round bin, reference to the shape of a car).

The major distinction between these two typologies is that in Loucks and Pederson (2010) there is no independent element for direction, but rather it can be derived from start and endpoints. Arguably, for route descriptions direction is a necessary component. However, it is important to also recognise that direction and path—while inseparable in a gesture—are in fact two distinct semantic categories. Ultimately, I draw from both these typologies since Loucks and Pederson (2010) is based on extensive research on how the world’s languages structure motion event descriptions whereas the typology of Holler and Wilkin (2009) is based on an analysis of speech and gesture patterns.

Taking a typical utterance one finds in the map task, it is possible to generalise and suggest that they frequently include seven semantic elements:

1. **POSITION (P)**: representation of the position of either a landmark or the particular point on the route. This can be represented using direct coordinates or it may be given as a position relative to some landmark or previously given part of the route. Speech examples include: “In the top right corner of the page”. Gesture examples include: pointing to a particular location in gesture space in order to depict a particular location on the map.
2. **GROUND (G)**: is the representation of an entity that is being used to anchor the position of the route or another entity. Speech examples include direct reference to landmarks. Gesture examples involve two-handed gestures where one hand

represents a landmark and the other hand represents another landmark or the route.

3. **FIGURE (F)**: is representation of a landmark that is being anchored to the position of another landmark (the route is not classed as a **FIGURE**). Although technically the route is usually acting as a figure, **FIGURE** was only marked when a landmark is being referred to using another as ground. Speech examples include: “the pyramid is in line with the youth hostel”. Gesture examples include two handed gestures where one hand represents a landmark and the other hand represents another landmark.
4. **DIRECTION (DIR)**: representation of the direction from point a to point b on the route. Speech examples include: “go up”, “go down”; Gesture examples include: manual movements in the direction the route is travelling, these include tracing gestures that move the index finger through space as if it is drawing the route.
5. **ORIENTATION (O)**: representation of orientation of two elements on the map. Therefore, orientation involves the route and one or more landmarks (as **GROUND** or **FIGURE**). Spoken examples include “the youth hostel is north of the lemon grove”. Gesture examples include two-handed gestures that depict two features on the map, one-handed gestures that depict the orientation of one object to another, and one handed gestures that depict the distance between two map features.
6. **MANNER (M)**: representation of route shape. Speech examples include: “do a loop round”, “diagonally”, “circle”; Gesture examples include: tracing a large curve with the index finger, moving the hand in an undulating manner. Tracing gestures, therefore, depict **MANNER** and **DIRECTION**.
7. **DISTANCE (DIS)**: representation of the distance between two elements on the map. Speech examples include “It’s two centimetres above it”. Gesture examples are similar to orientation except distance must be produced with two-handed gestures or one-handed gestures that depict a space between two elements. An example of this second kind of distance gesture would include the index finger and thumb being held apart while the rest of the hand is closed. In this example, the space

in between the thumb and index finger represents the distance between two elements.

In order to apply these features to the data, the data was tagged using ELAN. ELAN allows analysts to annotate video data on separate tiers, which are temporally tied to the video. Tiers can either be independent or dependent. Dependent tiers are embedded within and temporally tied to higher tiers. In terms of the semantic tagging for this study, first, gestures were broken into gesture units and gesture phrases on separate tiers. Then, tone units were delineated in relation to speech and gesture phrases. Both tone units and gesture phrases were tagged for semantic content. Taken together, these two units represent what has been called an idea unit (Kendon, 2004). Because the focus is on the semantic features, from this point on, it will be referred to as a semantic unit. Since overlap of tone units and gesture phrases is not 100%, a third tier was created that was dependent on the speech semantics tier. In this tier the semantics of the gesture phrases were aligned with the semantics of the tone units. This process was in order to ensure that information relating to speech and gesture could be extracted from ELAN.

Every tier was created for both givers and followers. The next two coding tags described were also dependent on the semantics in speech tier. However, they represent utterance features that exist on a higher level to semantic features.

### Frame of Reference

As described above, people invoke frames of reference when describing spatial events. The coordinate structure that is invoked by a frame of reference is realised through reference to a figure, which is the focussed element in the description, and one or more reference objects (often referred to as ground). Frames of reference can be broken into three distinct categories: Intrinsic, Relative and Global.

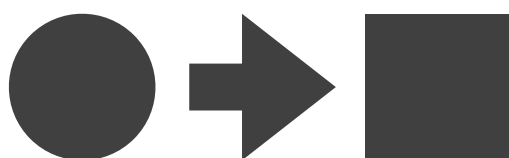


Figure 3.6

Turning to figure 3.6, the arrangement can be described from these three frames of reference:

- Intrinsic
  - Described arrangement of objects through inherent features
  - “The square is at the point of the arrow”
  - Figure: Square; Referent object: arrow
- Relative
  - Describe arrangement of objects using an egocentric perspective
  - “The circle is to the left of the arrow”
  - Figure: circle; primary reference object: arrow; secondary reference object: producer/comprehender perspective
- Absolute/Global
  - Describe arrangement of objects using an fixed coordinate structure
  - “The square is to the east of the circle”
  - Figure: square; primary reference object: circle; secondary (encompassing) reference object: cardinal coordinate structure (here the cardinal directions are not genuine, the main point is that the secondary reference object encompasses both the figure and primary reference object).
  - “The circle is behind the square in the line of shapes”
  - Figure: circle; primary reference object: square; secondary (encompassing) reference object: line of shapes

These three kinds of frame of reference are tied to the way participants in the map task describe the position of the route and its landmarks.

## Perspective

Perspective and frame of reference are fundamentally related. In the boat example described in above in section 3.1, there was a clear relationship between the frame of reference the individual was adopting and their gesture. However, in a situation like the map task, things are not always that simple. When people describe two dimensional objects like those found in the map task, they often depict those objects in gesture space. In such cases gesture space itself represents a portion of the map and gestures represent elements in analogous places on the map. From this perspective it could be argued that gesture space is acting as a secondary encompassing reference object and, as such, these gestures would be considered to invoke a global frame of reference. However, since the gestures are also usually tied to the perspective of the producer, they could be considered to invoke a relative frame of reference. It is for this reason that gestures are coded for perspective and speech is encoded for frame of reference. However, it is important to acknowledge that both exist on a higher level than speech and gesture alone.

When people gesture they tend to gesture from their own perspective, their interlocutor's perspective, or from a shared perspective. Additionally, Tversky et al. (2009) highlight the fact that interlocutors who are discussing maps gesture on the map so they can both see what is being described. In the map task, participants perform this process for themselves, out of sight of the person they are interacting with. The four perspectives that can be derived from this observation are as follows:

1. First Person (FP). The gesturer gestures in shared gesture space retaining their own perspective (i.e., left = left).
2. Second Person (SP). The gesturer gestures in shared gesture space adopting their interlocutor's perspective (i.e., left = right).
3. Shared (Sh). The gesturer gestures in shared gesture space sharing their interlocutor's perspective by turning their back to them.
4. Unshared (U). The gesturer gestures in unshared gesture space retaining their own perspective (i.e., they trace the route on the physical map).

### 3.3.4 Mechanical properties

Here, two mechanical properties are focussed on: number of hands used in gesture and handshape employed. Handshapes were analysed using a system largely derived from Stokoe notation (cf. Stokoe, 2005[1960]) and McNeill's 1992 system. Below only those handshapes that were realised in the map task are discussed.

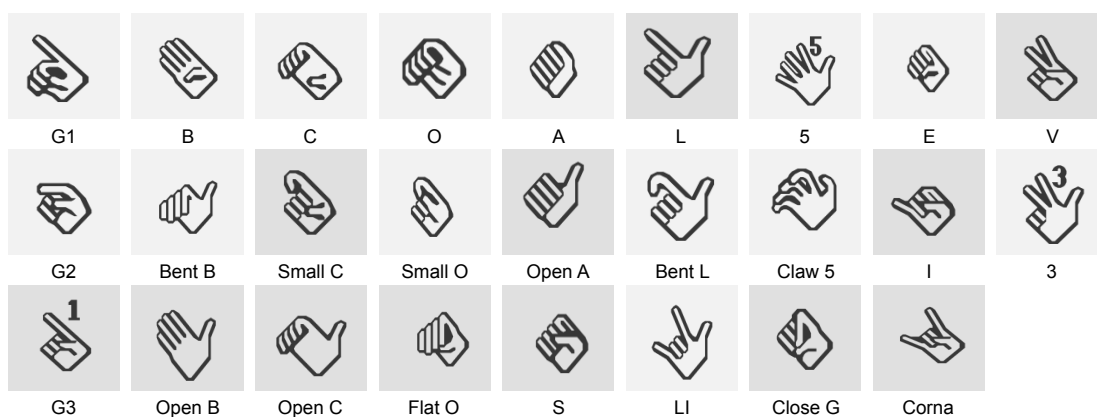


Figure 3.7: Handshapes used in the map task

G here is largely being used as a shorthand for pointing gesture and includes what would, following more traditional schemes (e.g., the one used by Stokoe), be referred to as D or 1.

### 3.3.5 Coding Example

The data was all coded using the transcription software ELAN. Figure 3.8 shows an example taken from the maptask analysis.



The screenshot displays the Elan software interface, which is used for video analysis and transcription. The top menu bar includes File, Edit, Annotation, Tier, Type, Search, View, Options, Window, and Help. Below the menu, there are three video thumbnails: a person in a red shirt, a person in a blue shirt, and a person writing on a whiteboard. To the right of the videos are navigation controls (play, stop, previous, next, first, last) and a selection box showing '00:12:18.624'.

Below the navigation controls is a 'Controls' panel with a volume slider set to 100 and three mute/solo sliders for 'Follower.mov', 'Giver.mov', and 'Map.mov'. Each slider has a 'Mute' and 'Solo' radio button. Below these are checkboxes for 'Selection Mode' and 'Loop Mode'.

The main area of the interface is a detailed Elan coding timeline. It shows a video frame at the top with a red selection bar. Below the frame, the timeline is divided into tiers. The 'Speech' tier contains the transcript: 'then up there and the it goes back down', 'so am I going over the top of the cliffs', 'no no it's gonna go up [and then this and then back down]', 'o[h: I see and then] okay', and 'so your gonna go <up to that> reference point we m'. The 'G\_game\_Coding' tier includes codes like '60.1.1.2.3.1 Inst' and '60.1.2.3 Ack'. The 'G\_gestureUnit' tier includes '60.1.2.2 Clarify'. The 'F\_gesturePhrase' tier includes '60.1.2.3 Ack'. The 'G\_gestureForm' tier includes 'LH: G' and 'LH: small C'. The timeline also shows time markers from 00:12:18.000 to 00:12:25.000.

Figure 3.8: Elan coding example 1

The three video feeds used for transcription and annotation are in the top left hand corner of figure 3.8. The left image depicts the giver, the top right image depicts the follower and the bottom right image depicts the map the follower is drawing on.

Moving on to the annotation stave, there is a timeline across the top, showing that this example begins at 12 minutes and 17 seconds and ends at 12 minutes and 25 seconds. The first tier, titled “G\_Speech”, contains the transcribed speech of the giver. All annotations relating to the Giver are labelled using the “G\_” prefix, whereas all relating to the follower are labelled using the “F\_” prefix. The next tier, titled “F\_Speech” contains the transcribed speech of the follower. Each annotation on these top two tiers is determined by the phonological structure of the utterance. Gesture was ignored in the segmentation of the G\_Speech and F\_Speech.

The next two tiers, titled “Game\_Coding”, contain the move coding outlined above. The giver’s first move is annotated as “60.1.1.2.3.1 Inst”. This annotations represents the fact that this move is embedded three levels and is the first pair part of the third project. Using the terminology of section 3.3.1, this project is the third E-project. The “1.2” represents that this move is embedded within the second pair part of the first embedded move, or first D-project. Finally, the “60.1” represents the fact that this is the sixtieth project in this maptask, or the sixtieth C-project. This move is also an instruct move, meaning that the giver is presenting new information that, if followed, will help the follower draw the map. Notice that the next annotated move is not coded as a check of “60.1.1.2.3.1 Inst”. If it was, it would be annotated “60.1.1.2.3.1.1.1 Check”, because checks are embedded within the moves they relate to. Instead, the follower’s check is annotated as “60.1.2.1 Check” which means that this represents a D-project check and is returning to a higher level of the task. The next move, “60.1.2.2 Clarify”, is a response to this check and is an attempt to clarify move “60.1”.

The next two tiers are titled “gestureUnit” and depict the gesture units as described in section 2.2.3 above. Each unit is delineated by rest points and therefore may contain multiple gesture phrases. Gesture phrases are depicted in the tiers titled “gesturePhrase”.

Figure 3.9 depicts coding below the level of the utterance and gesture phrase.

The screenshot displays the Elan software interface. At the top, there are menu options: File, Edit, Annotation, Tier, Type, Search, View, Options, Window, and Help. Below the menu is a navigation bar with buttons for Grid, Text, Subtitles, Lexicon, Comments, Recognizers, Metadata, and Controls. The main workspace is divided into several sections:

- Video Feeds:** Three small video windows showing participants in a meeting room.
- Playback Controls:** A set of buttons for navigating through the video (play, stop, previous, next, etc.) and a selection bar.
- Volume and Mute Controls:** Sliders and checkboxes for 'Follower.mov', 'Giver.mov', and 'Map.mov'. Each has a 'Mute' checkbox and a 'Solo' radio button. Volume sliders are set to 100.
- Timeline:** A horizontal axis showing time from 00:12:18.624 to 00:12:25.000. A red vertical line indicates the current playback position.
- Coding Tiers:** A list of annotated segments on the left side of the timeline, including:
  - F\_gesturePhrase [78]
  - G\_gestureForm [160]
  - F\_gestureForm [77]
  - G\_gestureSem [124]
  - G\_speechSem [161]
  - F\_gestureSem [60]
  - F\_speechSem [190]
  - G\_gestureSem[Spe] [181]
  - G\_perspective [181]
  - G\_FOR [140.4]
- Event Markers:** Vertical lines on the timeline indicating specific events like 'LH: G', 'RH: HP', 'Dir: M', 'Dir', 'Dir: M', 'Dir: G', 'Dis', 'FP', and 'NA'. A yellow box highlights a specific event: 'BT: 00:12:21.956, ET: 00:12:23.410 Dir'.

Figure 3.9: Elan coding example 2

The first tiers, titled “gestureForm”, depict the mechanical properties of the gesture. The first represents the form of the Giver’s gesture and is labelled “LH:G”, which describes what can be seen in the video still. “LH” refers to the fact that the giver is using his left hand and “G” refers to his handshape.

The next four tiers depict the semantic feature analysis of speech and gesture. The “G\_gestureSem” tier represents the semantic properties of the giver’s speech are determined clausally, so the giver’s speech “then up there and then it goes back down” is titled as being composed of two semantic units, both describing direction. Coniciding with this speech, the giver produces two gesture phrases depicting the direction and manner of the route. In other words, the gesture depicts more than the speech describes. Further, it can be seen that the giver continues to gesture while the follower produces her check. The semantic content of the check is coded as orientation and ground. The first relates to “going over the top” and the second relates to “of the cliffs”.

Figure 3.10 depicts the additional semantic analysis.

File Edit Annotation Tier Type Search View Options Window Help

Grid Text Subtitles Lexicon Comments Recognizers Metadata Controls

Volume: 100  
0 50 100  
Follower.mov

Selection: 00:00:00.000 - 00:00:00.000 0

Selection Mode Loop Mode

00:12:18.624

	2:17.000	00:12:18.000	00:12:19.000	00:12:20.000	00:12:21.000	00:12:22.000	00:12:23.000	00:12:24.000	00:12:25.000	00:--
F_speechSeri <sub>[181]</sub>	Dir; M	Dir; M			Dir; M	Dir; M		Dis		
G_gestureSemiSpe <sub>[181]</sub>										
G_perspective <sub>[181]</sub>	FP	FP			FP	FP		FP		
G_FOR <sub>[181]</sub>	NA	NA			NA	NA		NA		
G_around <sub>[181]</sub>	0	0			0	0		0		
G_motion <sub>[181]</sub>	1	1			1	1		1		
G_speechFramed <sub>[181]</sub>	0	0			0	0		0		
F_gestureSemiSpe <sub>[90]</sub>										
F_perspective <sub>[90]</sub>										
F_FOR <sub>[90]</sub>										
F_around <sub>[90]</sub>										
F_motion <sub>[90]</sub>										
F_speechFramed <sub>[90]</sub>										
	S									
	EGO									
	0									
	1									
	0									

Figure 3.10: Elan coding example 3

The first tier, titled “G\_gestureSem|SpeechSem” depicts the giver’s the semantic feature content of gesture by the semantic feature content of speech. This allows for the direct comparison speech and gesture. The next tier, titled “G\_perspective”, depicts the perspective of the giver’s gesture. In this example, it is coded as “FP” since the giver’s gesture is produced from the first person perspective. Next, “G\_FOR”, depicts the frame of reference invoked by the giver’s speech. Since there is nothing in the giver’s speech that can be coded as frame of reference, these units are coded as NA. The next three tiers represent whether the speech contains the word “around”, describe motion, or whether the gesture is speech framed.

Comparing the giver’s tiers to the follower’s tiers, the follower’s gesture is from a shared perspective, because she turns so that her right and the giver’s right are the same. The follower’s speech describes the orientation from an egocentric perspective, due to the use of the phrase “over the top”.

The ELAN annotations were exported as .csv files for analysis.

### 3.3.6 Research Questions

This chapter has outlined how the data generated using the map task were broken down for analysis. In this thesis, data relating to route reduplication, and gesture repetitions are not presented. These present avenues for future research.

The next chapter presents the analysis of the map task data. Its aim is to answer the following research question:

- What is the effect of gesture on production of semantic units? More specifically:
  - How is information conveyed through semantic features distributed across speech and gesture?
  - From what perspective is gesture produced and what effect does gesture have on it?
  - What is the effect of gesture on frame of reference?

- Are gestures more likely to occur with different move types?
- What is the effect of gesture on structure of joint projects?

These questions aim to explore the relationship between what is commonly considered to be related to semantics and the role of gesture. The general assumption is that the more closely related speech and gesture are the more likely it is that gesture is part of the ostensive inferential model of communication and thus provide evidence for a composite signal view of language and gesture.





# Chapter 4

## An analysis of Gesture in the map task

### 4.1 Introduction

This analysis explores the various ways in which gestures are used during the map task and the effect they have on the interactive sequence. To do so it focusses on speech and gesture from several distinct vantage points.

- As part of utterance production
  - The types of gesture found during the map task
  - The semantic meaning of the utterance as is it made manifest through speech and gesture (i.e., what elements of the referent are highlighted by the components of the utterance)
  - The frame of reference underlying each utterance
  - The perspective taken by the utterer
  - Descriptions of motion
  - Utterance complexity
  - Lexical affiliates (a case study of the word “around”)
- As part of the interactive sequence

- Move types and gesture
- Gesture’s effect on the project level

Most of these categories were outlined in chapter 3, however some developed as a result of the salient phenomena that emerged during the map task. Where relevant throughout the analysis, detailed examples will be provided.

## 4.2 Production

### 4.2.1 Gesture in the map task

Following the coding scheme outlined in the previous chapter, 2021 semantic units have been coded, with 39% of the semantic units containing information conveyed through gesture (presented in table 4.1). In other words 39% of semantic units contained representational gestures. The units counted as containing gesture excluded beat gestures (McNeill, 1992) and, what are often called, interactive or interpersonal gestures (Kendon, 2004). Such gestures are related to the discourse context and not, necessarily, to the semantic content of the utterance.

<b>Speech only</b>	<b>With gesture</b>
1240 (61%)	781 (39%)

Table 4.1: Semantic units containing speech only and gesture

Considering the participants are describing space, this might seem like a relatively low proportion. However, it is important to acknowledge that 781 gesture tokens still represents a fairly large corpus. For comparison, the Bielefield Speech-and-Gesture-Alignment (SaGA) corpus is 280 minutes and 4961 gestures (Lücking et al., 2013). Additionally, one of the purposes of this analysis is to explore the environment(s) in which gestures do not appear as well those in which they do. Ultimately, if we are to develop a theory of gesture from a pragmatic perspective, then it is crucial to not only explain why people gesture, but also why they do not.

Throughout this analysis, statistical evidence is provided using a linear mixed effect analysis (Bates et al., 2015). However, all analyses should be taken with caution due to

Task	Time	Participant	Semantic Units	Gesture
MT01	52.36	P01	405	0.24
		P02	161	0.12
MT02	9.44	P03	132	0.74
		P04	35	0.46
MT03	5.27	P05	50	0.96
		P06	18	0.17
MT04	14.28	P07	195	0.56
		P08	98	0.57
MT05	4.25	P09	27	0.33
		P10	16	0.44
MT06	9.08	P11	131	0.23
		P12	43	0.56
MT07	9.25	P13	81	0.09
		P14	16	0.38
MT08	38.58	P15	352	0.59
		P16	261	0.17
Totals	92.18		2021	0.39

Table 4.2: Distribution of data by participant. This table shows the number of semantic units produced by each participants and the proportion of those units that contain gesture

the low number of participants involved in the study. Echoing Winter’s (2015) concerns, it is important not to use the power that results from a high number of observations to hide what is determined on the basis of participant variation. Further, a fuller description of the model fitting procedure can be found in chapter 6 and the appendix. Additionally, this first section of this analysis focusses only on the semantic content of utterances, ignoring the move that it is part of. Move type will be focussed on the section 4.3.1 below. For this reason, the next section will explore the effect of participants and (briefly) move types on the presence of gesture.

## 4.2.2 Participants and Move Types

### Individual Variation

Task specific variation was high. Table 4.2 shows the variation in terms of time, number of semantic units and the proportion of semantic units containing gesture.

Furthermore, it was found that the participant producing the semantic unit has a significant effect on whether gesture is produced as part of the unit ( $\chi^2(17) = 428.88, p = <$

0.001)<sup>1</sup>.

### The effect of move type

Another potential for variation is the way gesture occurs with the type of move. If we only include those move types that account for more than 5% of the corpus<sup>2</sup>, the distribution is represented in figure 4.1.

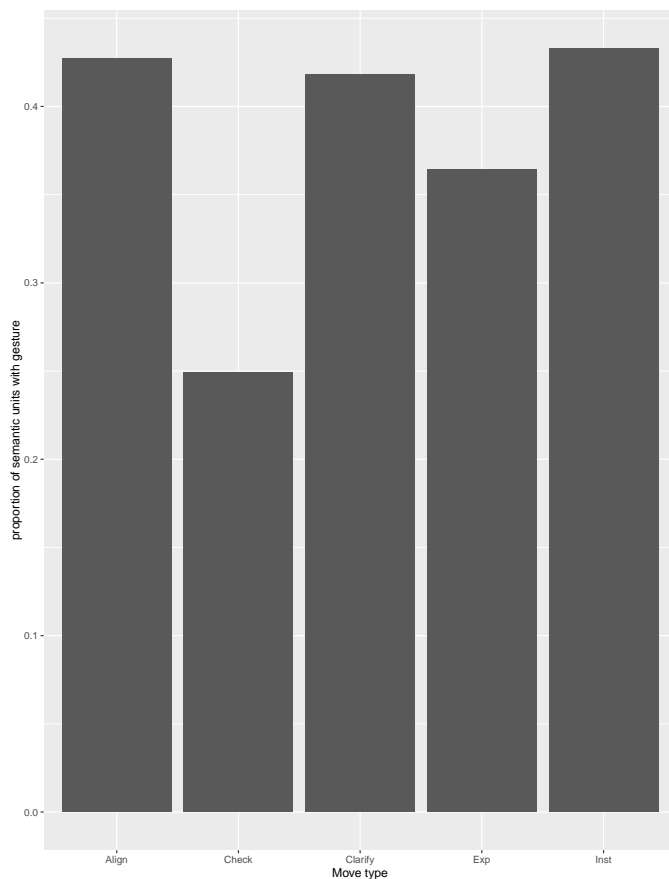


Figure 4.1: Gesture by Move Type

From graph 4.1 it is clear that there are fewer gestures with checks than other move types. What's more, like participant, move type has a significant effect on the type

<sup>1</sup>In this analysis linear mixed effects models are fitted to the data and p-values are generated using a likelihood ratio test. This test compares the model of the outcome (in this case the incidence of gesture) with a predictor (in this case participants) to a zero intercept model. The zero intercept model takes the average value of the outcome. In other words, this analysis compares the average incidence of gesture to the incidence of gesture when constrained by each level of the predictor. A full description of this method is described in chapter 6 below.

<sup>2</sup>5% was chosen because this represented all moves that have more than 100 tokens. Further, this captures all moves that are likely to involve spatial descriptions

of gesture found in the map task ( $\chi^2(13) = 29.317, p = 0.001107$ ). A more detailed analysis of the effect of move type on gesture production will be presented in section 4.3.1 below.

These two findings highlight the lack of regularity in the corpus. For this reason, all analyses employ linear mixed effects models, taking move type and participant as random intercepts and (where possible) slopes.

### 4.2.3 What kind of gestures are found in the map task?

This section provides an overview of the mechanical properties of gestures produced during the map task. These will be discussed in terms of hand shape and number of hands used in gestures during the map task. These are not limited to only those gestures conveying semantic information but to all manual gestures.

One-Handed	Two-Handed
874 (77 %)	259 (23%)

Table 4.3: Number of hands used during gesture production

Table 4.3 demonstrates that one-handed gestures were far more common in the map task. On occasion, participants use two hands, each representing the different location of a feature on the map. These features could be two landmarks or a landmark and the route. In these cases, gestures typically depict what Enfield (2009; 2004) has called symmetry-dominance constructions (borrowing the term from sign language phonology). While this behaviour would be more explicit in depicting the route or landmark position, the fact that participants tended to adopt one-handed gestures suggests that this was a relatively uncommon strategy.

Hand Shape	Number (%)
G	478 (34%)
B	259 (19%)
HP	172 (12%)
5	111 (8%)
open B	67 (5%)

small C	66 (5%)
bent B	48 (3%)
small O	29 (2%)
interlaced	25 (2%)
L	21 (2%)
5 claw	17 (1%)
Flat O	16 (1%)
O	16 (1%)
C	13 (1%)
E	11 (1%)
Folded Arms	10 (1%)
flat C	6 (<1%)
Open A	4 (<1%)
Open C	4 (<1%)
S	4 (<1%)
9	4 (<1%)
A	3 (<1%)
bent L	3 (<1%)
close G	3 (<1%)
I	3 (<1%)
V	3 (<1%)
3	2 (<1%)
cornu	1 (<1%)
LI	1 (<1%)
open H	1 (<1%)

Table 4.4: Handshapes used during Map Task

Table 4.4 shows the distribution of the different handshapes across the task. There are some additional labels that were not addressed in the original coding scheme. “HP” refers to those gestures in which the producer was Holding the Pen. In these instances, the pen

was being employed in an identical manner to an index finger in a pointing or tracing gesture. “Interlaced” refers to gestures produced while the fingers were interlaced and “Folded Arms” refers to gestures produced while the arms were folded. Although these two categories may not be, strictly speaking, manual gestures, they were often used in a similar manner. For example, a producer with folded arms could move their arms together, pushing their right elbow away from their body whilst moving their left elbow towards it. This movement could be used to refer to a rightward directed route shape, or the relative position of a landmark.

Overall, there is a clear preference in the data for pointing gestures (G, HP, L, bent L, I, corna) or gestures involving a flat hand (B, 5, open B, bent B). Both of these groups of handshapes can be employed during tracing and positioning objects in gesture space. Pointing gestures can be additionally employed for directly pointing at entities within the map or the producer’s interlocutor.

#### 4.2.4 What is referred to (and in what medium)?

This section focusses on how different semantic features are distributed across speech and gesture. There are two distinct hypotheses relating to the two pragmatic views of gesture production:

1. If gestures are being used to help speakers think about talking about space, then gesture should tend to mimic speech
2. If gestures are being used communicatively, then gesture should present information not found in speech

Table 4.5 contains the distribution of semantic information across the map task.

	<b>P</b>	<b>G</b>	<b>F</b>	<b>Dir</b>	<b>O</b>	<b>M</b>	<b>Dis</b>
S	388 (19%)	1093 (54%)	91 (4%)	583 (26%)	640 (32%)	606 (30 %)	316 (16 %)
G	220 (10%)	63 (3%)	10 (<1%)	436 (21%)	116 (5%)	495 (25%)	67 (3 %)
B	62	33	4	194	58	234	32
S + G	16%	3%	4%	33%	11%	39%	10%
G + S	28%	52%	40%	45%	50%	47%	48%

Table 4.5: Distribution of Semantic information across the map task based on 2021 semantic units. The number of semantic units and their percentages across the whole corpus is presented for each semantic category in both speech (S) and gesture (G). The columns represent each of the semantic features discussed in chapter 3: POSITION; GROUND; FIGURE; DIRECTION; ORIENTATION; MANNER; and DISTANCE. The row labelled “B” contains the tokens where the column’s semantic category is presented in both speech and gesture. Those tokens in B also appear in the S and G rows. “S + G” presents the percentage of speech tokens that also contain gesture (irrespective of semantic category) and “G + S” presents the percentage of gesture tokens containing speech (regardless of semantic category). Therefore, taking M, this data shows that 39% of speech describing MANNER includes gesture depicting MANNER, which also means that 61% of speech describes MANNER without gesture. The column depicting POSITION is represented diagrammatically in figure 4.2

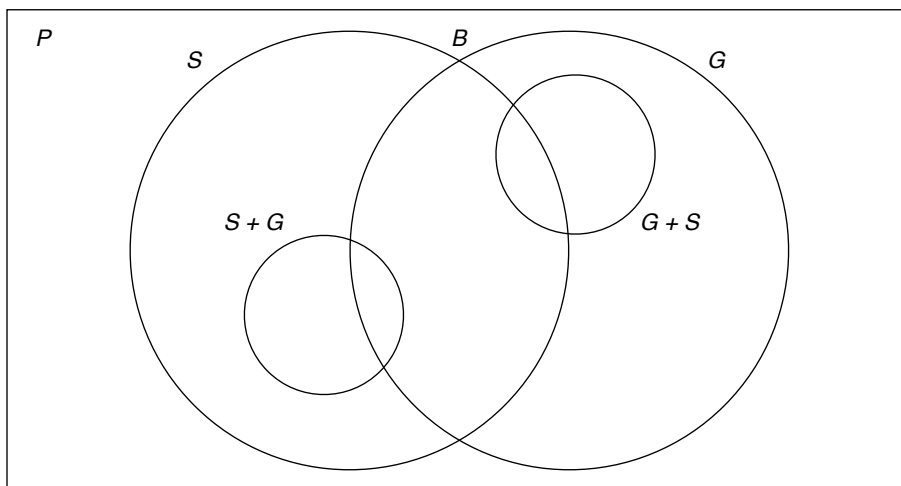


Figure 4.2: Representation of Position information in table 4.5





Table 4.5 demonstrates that while information conveyed through speech is more prevalent in the corpus than information conveyed through gesture, there is a different distribution of semantic elements in the two modalities. In 54% of the corpus, GROUND is represented through speech vs only 3% in gesture. This discrepancy follows from the fact that since GROUND relates to the explicit labelling of an object that can be used to describe the relative position of other objects (including the route), then it can be more easily expressed through speech (through noun phrases, for example) than through gesture. For gesture to perform this function it would have to be represented using a two-handed (“symmetry-dominance”) construction with the non-dominant hand depicting the GROUND. As has already been shown, two-handed gestures were less common than one-handed gestures in the corpus.

POSITION is similar to GROUND because they can both be used to highlight a particular point on the map. However, gesture is more prevalent in the representation of POSITION than it was for GROUND. This is likely due to the fact that position may be depicted gesturally by pointing to a particular point in gesture space (McNeill, 1992). In this sense then, POSITION is relatively easy to refer to using gesture since a particular position in gesture space (e.g., the top right) correlates easily with a particular position on the map (e.g., the top right corner of the map). However, in order to depict a particular entity, gesture would have to represent that entity, not just in terms of POSITION, but also as existing in that POSITION. This would involve modelling the entity in gesture space (Enfield, 2009b). For example, by using a fist (e.g., A or S handshape) or using a handshape that depicts something as being of a particular size (e.g., (small) C or (small) O handshape), a producer not only picks out a referent’s position in space (which may or may not be meaningful) but they also depict the referent as having physical properties (e.g., size) that take up a part of gesture space. Ultimately, it seems that there are clear reasons why the discrepancy between speech and gesture in the description/depiction of GROUND is not as large in the description/depiction of POSITION.

The next semantic element in 4.5 that represents a large discrepancy between speech and gesture is ORIENTATION. This was an unexpected finding since studies have shown that gesture is a good resource for people describing the relative position of two ob-

jects (cf. Holler and Beattie, 2003a). ORIENTATION was tagged only when it is explicitly referred to either through speech, gesture, or both. For speech, it was tagged when it describes the POSITION of something (e.g., route shape or landmark) relative to something else, such that the relative position of those two things is made explicit (e.g., by using words like “above” or phrases like “on the right of”). For gesture, this relationship can only be realised explicitly if a two-handed gesture or a one-handed gesture, where the hand depicts two separate points (e.g., one using a small C handshape), is employed. As we have seen above, most gestures were one-handed and there was a relatively low incidence of small C handshapes (5%), which suggests that gestures typically either depict points in gesture space, objects in gesture space, or the route. Therefore, while ORIENTATION may be derivable from the sequential placement of a route relative to some previously gestured position, it is not tagged as ORIENTATION since it is not made explicit. Furthermore, and far more prevalent in the data, is that ORIENTATION is not explicitly marked in gesture but it is derived by enrichment of a gesture relative to the content of speech. An example<sup>3</sup> can help highlight this point.

1	F:	okay (0.5) s::o:, (0.7) er::::m well I	
2	F:	guess I'll just alter my peak=[so	
3	F:	(.)]1 [I'm not getting does the	
4	F:	peak go over]2 [the top of the:	
5	F:	(0.6) tri=your pyramid	
6	G:	yeah]3	

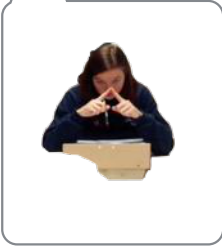


Figure 4.3: Example of orientation

<sup>3</sup>Throughout this analysis examples will be used to help illuminate the distribution highlighted in the statistics.

Before presenting the analysis, there are two important things to point out regarding the extract in figure 4.3. First, it depicts a sequence of gestures produced by the follower. Second, these gestures accompany a CHECK. This example, therefore, demonstrates the importance of trying to maximise the different types of data present in the corpus.

In figure 4.3, F's turn begins with an explanation of a change F is going to make to the peak of the route she is drawing as it goes over the top of the pyramid (see figures 3.4 and 3.5 for the maps). This change is the result of a clarification G made to his description of the route. The elongation and pauses found in this first turn, project that what F is saying is going to be a potential trouble source as G and F complete the task. This explanation is latched to a discourse marker on line 2 ("so") which introduces the CHECK. Concurrent with this marker is a gesture "displaying communicative action" (depicted in the first image) involving the index finger pointing upward (Streeck, 2009). This gesture is similar to a beat in that it regulates talk, however it operates at the discourse level mirroring the co-temporal "so". In doing so, both speech and gesture mark a shift from the explanation of the problem to the beginning of the CHECK. The CHECK is prefaced with an assessment of F's epistemic status ("I'm not getting" on line 3), explicitly orientating to her lack of understanding (Heritage, 2012). This framing, up to the word "peak", is concurrent with the preparation phase of the gesture, which involves F moving her hand, holding her pen as if drawing, from a lateral position towards a counter-lateral one. In this case, the movement goes from her right to her left. Importantly, at this point the movement is not meaningful, it is not depicting a trace. This movement stops in line with her right eye and its final position is depicted in the top image tied to the second gesture of the extract in figure 4.3. The word "peak" is realised with a pre-stroke hold, and the stroke is concurrent with "go over" on line 4. The temporal placement of the gesture stroke in relation to the speech is timed so that the trace formed by the gesture depicts the route described by the phrase "go over". This time then, the movement does produce a trace. The images related to this gesture depict the start and end point of the stroke.

As explained above, the word "over" is tagged for ORIENTATION, however the concurrent gesture is not. That is because there is nothing inherent in the gesture on its

own that explicitly depicts the path as being oriented to anything, the gesture depicts the DIRECTION and MANNER of the route. The gesture, which depicts a rightward moving route that is curved, must be enriched using the ORIENTATION in the speech so that the fact that this route is above the pyramid can be derived. What's more, although the word "peak" was tagged for MANNER, it is underspecific, because it only partially describing the shape of the route (i.e., as having a peak). It is the gesture that explicitly provides information relating to the curvature of the route. Taken together, not only does the gesture present information that enriches the meaning of the spoken component, but the speech presents information that situates the gesture in gesture space. As suggested by Kelly et al. (1999, p. 583) "speech is context for gesture just as gesture is context for speech". The curve in the route depicted by the gesture is now a curve over the pyramid, and therefore the space under the curve is now occupied by the pyramid—even though it has never been explicitly stated (in neither gesture nor speech) that this is the case. Therefore, the ORIENTATION of the gesture can only be understood as a product of the interaction between the gesture and speech. This is because they have been produced to be both phonologically and semantically synchronous (McNeill, 1992, pp. 26–28). The reasons these two utterance components are taken together is captured by Enfield's (2009, pp. 16–17) *contextual association* and *unified utterance-meaning* heuristics, which specify that contextually associated signs should be taken together and processed as a single whole.

The final gesture (lines 5 & 6) is a modelled depiction of the triangular shape of the pyramid (Enfield, 2009b; Streeck, 2008). Since F has an old temple on her map where G's pyramid is, it could be that this gesture is produced to help F retrieve the name of the analogous landmark on G's map (Yap et al., 2011; Krauss, Chen, and Gottesmann, 2000). This idea is further emphasised by the self initiated, self repair on line 5, in which F begins to say what sounds like "triangle" (a related term to pyramid and the description of the shape F is producing) (Schegloff, Jefferson, and Sacks, 1977). Importantly, the gesture depicting the pyramid is not placed in a position so that the route shape depicted in the gesture on lines 3 & 4 is oriented relative to it. In other words, the coordinates of gesture space adopted by this final gesture are not meaningful. If they were, the

triangular gesture would be placed further to F's right. These gestures are sequential and synchronised with the phonological realisation of speech rather than the, higher-level, semantic meaning of the utterance. This adds credence to the fact that tracing gestures should not be automatically tagged as depicting ORIENTATION.

Another important feature to point out regarding example 4.3 is that it highlights the aspect of the analysis to be presented throughout this chapter. The CHECK, presented on lines 3-5 would have been tagged as a single speech unit consisting of MANNER, ORIENTATION, and GROUND. The gesture accompanying the speech would have been tagged as two gestures, one depicting MANNER and DIRECTION and the other depicting GROUND. In the analysis, gesture units are treated in relation to speech, so that this gesture would have been tagged as depicting MANNER, DIRECTION and GROUND, when in reality the underlying semantic unit is decomposed of two gesture phrases.

Figure 4.3 is just one example of ORIENTATION in the corpus. Table 4.6 shows the distribution of other elements relative to ORIENTATION throughout the entire corpus.

	<b>P</b>	<b>G</b>	<b>F</b>	<b>Dir</b>	<b>M</b>	<b>Dis</b>
<i>by ORIENTATION N = 698</i>						
<i>Speech</i>	92 (13%)	545 (78%)	36 (5%)	134 (19%)	112 (16%)	151 (25%)
<i>Gesture</i>	79 (11%)	62 (9%)	10 (1%)	113 (16%)	119 (17%)	34 (5%)
<i>Both</i>	17	32	4	44	44	17
<i>by ORIENTATION with gesture N = 250</i>						
<i>Speech</i>	33 (13%)	162 (65%)	10 (4%)	50 (20%)	50 (20%)	44 (18%)
<i>Gesture</i>	76 (30%)	62 (25%)	10 (4%)	113 (45%)	119 (48%)	30 (12%)
<i>Both</i>	17	32	4	44	44	15
<i>by ORIENTATION in gesture N = 116</i>						
<i>Speech</i>	25 (22%)	60 (53%)	9 (8%)	8 (6%)	15 (13%)	28 (20%)
<i>Gesture</i>	30 (26%)	61 (53%)	10 (9%)	26 (22%)	30 (26%)	23 (20%)
<i>Both</i>	13	31	4	7	12	14
<i>by ORIENTATION without gesture N = 448</i>						
<i>Speech</i>	59 (13%)	383 (86%)	26 (6%)	85 (19%)	62 (14%)	105 (24%)

Table 4.6: Distribution of elements relative to Orientation

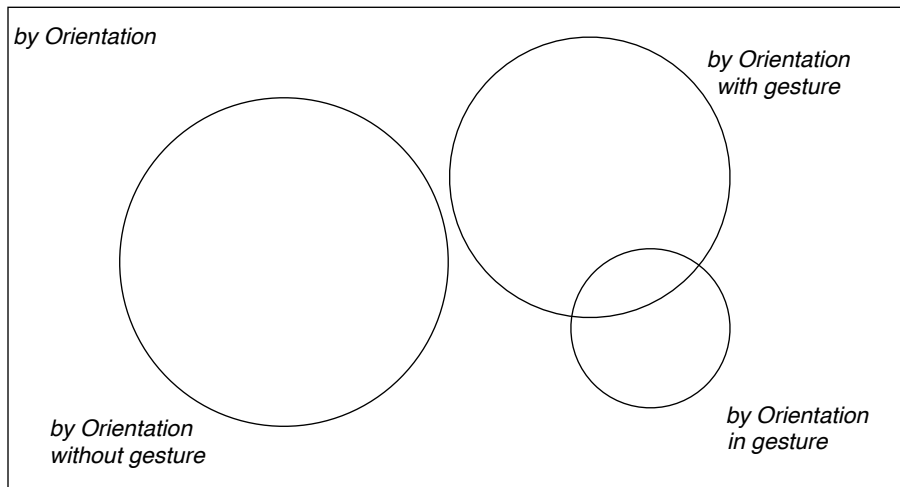


Figure 4.4: Representation of Orientation information in table 4.6

Table 4.6 shows the relationship between the description of ORIENTATION in speech and the depiction of ORIENTATION in gesture. The relationship between the different subtables is presented in 4.4. First, as shown in table 4.6 only 11% of units containing ORIENTATION in speech also contain ORIENTATION in gesture. However, due to the relatively small incidence of ORIENTATION in gesture (116 tokens in the corpus), this does not mean that this relationship goes in the other direction. 50% of semantic units containing ORIENTATION in gesture are accompanied by ORIENTATION in speech. This suggests that the presence of ORIENTATION information in gesture may be motivated by its presence in speech. This relationship is statistically significant, where ORIENTATION in gesture is predicted by ORIENTATION in speech ( $\chi^2(8) = 6.9234, p = 0.008507$ ) increasing its presence ( $\beta = 0.9119 \pm 0.2833(SE), z = 3.218, p = 0.00129$ ). However, it remains the case that 50% of ORIENTATION gestures are not accompanied by ORIENTATION in speech.

Going back to the first subtable of table 4.6, which limits the corpus to just those units containing ORIENTATION in either speech or gesture, there is an increase in the incidence of GROUND conveyed through speech: (78%) compared to the full corpus (54%).

What's more, the incidence of GROUND in speech decreases to 65% when only those units containing gesture are considered (as shown in the second subtable). This trend continues to 53% when only ORIENTATION in gesture is considered. However, in the final subtable, which shows the ORIENTATION in speech, excluding any contribution of gesture, the incidence of GROUND is increased to 86%. It seems, then, that the presence of gesture seems to have a negative effect on the presence of GROUND in speech. Analysed statistically, although there was not a main effect of gesture on GROUND in speech ( $\chi^2(8) = 2.4907, p = 0.1145$ ) paired samples t-tests do show that gesture is significantly reducing GROUND in speech ( $\beta = -0.7944 \pm 0.3871(SE), z = -2.052, p = 0.0401$ )<sup>4</sup>.

Another trend that emerges in table 4.6 is that the incidence of DIRECTION and MANNER information in gesture increases when only those units containing gesture are considered. However, this is generally true of the whole corpus. Generally, gesture seems to represent MANNER and DIRECTION more than any other category. However, when compared to the full corpus, in 4.6 there is a reduction in MANNER (from 30% to 17%) and DIRECTION (from 21% to 16%) gestures. This reduction is also occurs in the amount of MANNER (from 30% to 16%) and DIRECTION (from 26% to 19%) present in speech. This is something that will be discussed further below.

Focussing on the internal differences within table 4.6, as is the case with speech, one of the biggest differences when comparing the ORIENTATION with gesture depicted in the second subtable and the general distribution of semantic information in gesture can be found in GROUND. The incidence of GROUND in gesture increases further when the corpus is limited to only those units that contain ORIENTATION in gesture. This effect is statistically significant ( $\chi^2(8) = 23.844, p = < 0.001$ ). Above, it was stated that ORIENTATION is linked to GROUND, because ORIENTATION describes the place of an object (usually the route or landmark) relative to another. The interesting thing is that while presentation of ORIENTATION in gesture is tied to the presentation of ORIENTATION in speech, they actually seem to appear in a distribution where the presence of gesture leads to a reduction in GROUND in speech and the presence of ORIENTATION in gesture leads to an increase in the presence of GROUND in gesture. This suggests that ORIENTATION gestures (50% of which do not occur with ORIENTATION in speech) are being used

---

<sup>4</sup>Paired-samples t-tests were carried out using lsmeans in R.

in two distinct contexts. In one context they mirror the information in speech, but in the other they complement speech.

Table 4.6 also strengthens the claim that ORIENTATION in speech can help guide the enrichment of DIRECTION and MANNER information in gesture. In other words, speech can provide a description of the ORIENTATION relative to a landmark and the gesture provides the DIRECTION and MANNER information of that particular part of the route. This idea can be derived from the fact that when the corpus is limited to only those units containing ORIENTATION and gesture, the incidence of DIRECTION and MANNER in gesture are 45% and 48% respectively. However, when only gestures that contain ORIENTATION are considered the incidence of DIRECTION and MANNER drops to 22% and 26%. In other words, DIRECTION and MANNER are regularly conveyed through separate gestures to ORIENTATION. Exploring these findings statistically, ORIENTATION in gesture significantly affects the presence of MANNER in gesture ( $\chi^2(8) = 4.2287, p = 0.03975$ ), resulting in a reduction in gesture ( $\beta = -1.5613 \pm 0.8758$ ). However, the relationship between ORIENTATION in gesture and DIRECTION in gesture is not significant ( $\chi^2(8) = 3.0513, p = 0.08067$ ), however this is still based on a reduction ( $\beta = -1.3352 \pm 0.8513$ ). These results suggest that gesture is performing several different functions. It may mirror speech, when speech and gesture both convey ORIENTATION or it may complement it when speech describes ORIENTATION and gesture depicts DIRECTION and MANNER.

One of the things highlighted by the distribution of semantic information relative to ORIENTATION is the fact that GROUND in speech occurs in an environment that does not contain gesture. Therefore, it is worth exploring this finding in the corpus generally. Returning to table 4.5, DIRECTION and MANNER information in gesture have the highest incidence of any gestured information (21% and 25% respectively). This suggests that gestures are used most in the map task to represent the route.



	<b>P</b>	<b>G</b>	<b>F</b>	<b>Dir</b>	<b>O</b>	<b>M</b>	<b>Dis</b>
<i>Semantic units containing gesture (N = 781)</i>							
S	98 (13%)	308 (39%)	30 (4%)	236 (30%)	172 (25%)	268 (34%)	74 (10%)
G	220 (28%)	63 (8%)	10 (1%)	436 (56%)	116 (25%)	495 (63%)	67 (6%)
B	62	33	4	194	58	234	32
<i>Semantic units not containing gesture (N = 1240)</i>							
S	290 (23%)	785 (63%)	63 (5%)	302 (24%)	448 (36%)	338 (27%)	302 (20%)

Table 4.7: Distribution of Semantic information with and without gesture.

Table 4.7 reiterates what has been highlighted already. When only those semantic units containing gesture are considered there is a reduction in GROUND in speech compared to when gesture is not present (from 63% to 39%). Furthermore, there is an increase in DIRECTION (24% to 30%) and MANNER (27% to 34%). This decrease of GROUND in speech in the environment of gesture is of a much greater magnitude than the increase in DIRECTION and MANNER in speech (-24% versus +6% & +7%, respectively). This suggests that DIRECTION and MANNER information in speech are far less affected by the presence of gesture than GROUND information in speech. What’s more, bearing in mind that 55% of DIRECTION gestures and 53% of MANNER gesture are not accompanied by speech containing analogous semantic features, it is not necessarily the case that gesture tends to simply mirror the speech it accompanies.

Exploring these features, the presence of gesture is tied to a reduction in the presentation of GROUND in speech ( $\chi^2(8) = 10.155, p = 0.001439$ ) with gesture reducing GROUND in speech ( $\beta = 0.9042 \pm 0.2304(SE), z = -3.925, p = < 0.001$ ). This effect goes the other way with the presence of GROUND in speech affecting gesture ( $\chi^2(8) = 10.832, p = < 0.001$ ). Analysing the effect of gesture on MANNER and DIRECTION in speech show that while gesture significantly predicts the presence of DIRECTION in speech ( $\chi^2(8) = 6.7629, p = 0.009307$ ) resulting in an increase ( $\beta = 0.05503 \pm 0.2157(SE), z = 2.551, p = 0.0107$ ) this is not the case for MANNER ( $\chi^2(8) = 0.2781, p = 0.5979$ ). This suggest that participants in the map task are more likely than not to use gesture when they talk about DIRECTION, but not necessarily when they de-

scribe MANNER.

Turning now to the relationship between MANNER and DIRECTION in speech and GROUND in speech, MANNER significantly affects GROUND in speech ( $\chi^2(8) = 9.698, p = 0.001845$ ) leading to a reduction in GROUND in speech ( $\beta = -1.0564 \pm 0.2158, z = -4.896, p = < 0.001$ ) and so did DIRECTION ( $\chi^2(8) = 4.1982, p = 0.04047$ ) also reducing the presence of GROUND in speech ( $\beta = 0.4583 \pm 0.2042, z = -2.244, p = 0.0248$ ). If explored the other way round, GROUND in speech significantly affects MANNER in speech ( $\chi^2(6) = 18.152, p = < 0.001$ ), reducing the incidence of MANNER in speech ( $\beta = -1.2437 \pm 0.2009, z = -6.191, p = < 0.001$ ). However, GROUND does not significantly affect DIRECTION in speech ( $\chi^2(8) = 2.2545, p = 0.1332$ ). These findings suggest that although participants are more likely to talk about GROUND without DIRECTION than they are to talk about both together, they are no less likely to DIRECTION without GROUND than they are with it. However, they do not talk about MANNER and GROUND together.

Looking more closely at the relationship between MANNER and DIRECTION in speech and gesture, they are all significantly correlated with each other. MANNER in speech predicts MANNER ( $\chi^2(8) = 13.393, p = < 0.001$ ) and DIRECTION ( $\chi^2(8) = 5.1231, p = 0.02361$ ) in gesture. Similarly, DIRECTION in speech predicts MANNER ( $\chi^2(8) = 8.5958, p = 0.003369$ ) and DIRECTION ( $\chi^2(8) = 11.647, p = < 0.001$ ) in gesture. Interestingly, this goes the other way, with MANNER in gesture predicting DIRECTION ( $\chi^2(8) = 10.361, p = 0.001287$ ) and MANNER ( $\chi^2(8) = 9.1228, p = 0.002524$ ) in speech. What's more, DIRECTION in gesture predicts DIRECTION ( $\chi^2(8) = 12.959, p = < 0.001$ ) in speech. However, DIRECTION in gesture does not significantly predict MANNER in speech ( $\chi^2(8) = 2.9423, p = 0.08629$ ) even though the estimated parameters do demonstrate an increase in MANNER in speech ( $\beta = 0.6894 \pm 0.3576, z = 1.928, p = 0.0539$ ). Although this result is almost significant it does suggest a potentially interesting feature of the distribution.

However, when the same comparisons are made within each modality, it turns out that MANNER in speech does not predict DIRECTION in speech ( $\chi^2(8) = 0.08556, p = 0.355$ ) and nor does DIRECTION in speech predict MANNER in speech ( $\chi^2(8) = 0.1905, p = 0.6625$ ). In gesture the picture is different with MANNER significantly predicting DIRECTION ( $\chi^2(8) = 33.473, p = < 0.001$ ) and DIRECTION predicts MANNER ( $\chi^2(8) =$

37.482,  $p = < 0.001$ ). This suggests that in gesture MANNER and DIRECTION regularly act as a single unit. In speech, however, they seem to occur in different environments.

To summarise these results, gesture and GROUND in speech appear to be in a distribution where the presence of one is tied to the reduction of the other. DIRECTION in speech and gesture are in the opposite relationship where they regularly co-occur. In other words, participants are more likely to gesture than not when they are describing DIRECTION. However, this is not the case for MANNER in speech. This suggests that MANNER is not unlikely to occur without gesture. The relationship between GROUND in speech and MANNER in speech is similar to the relationship between gesture and GROUND in speech. However, it is not the same case for the relationship between DIRECTION and GROUND in speech. The suggestion is that there is less GROUND information within the presence of DIRECTION information than elsewhere, but that there is not less DIRECTION information in the presence of GROUND than elsewhere. Therefore, it is possible that participants are describing DIRECTION relative to the landmarks, but they do not do this with MANNER.

Turning to the relationship between MANNER and DIRECTION in gesture and speech, participants are more likely to be gesturing about both DIRECTION and MANNER when they talk about DIRECTION and MANNER. Suggesting that gestures often represent the information described in speech. Furthermore, they are more likely to be talking about DIRECTION when their gestures depict DIRECTION and MANNER. However, only MANNER in gesture predicts MANNER in speech and thus participants are no more likely to be talking about MANNER when they are gesturing about DIRECTION than elsewhere. Finally, while DIRECTION and MANNER in gesture regularly co-occur, this is not the same for MANNER and DIRECTION in speech which neither co-occur nor appear to correlate negatively.

These findings suggest that there are potentially four types of spatial descriptions used in the map task. First, participants describe GROUND without directly referring to the route. A good example of this is where participants name landmarks sequentially. In other words, the shape of the route is described by specifying which landmark will come next. Second, participants describe GROUND and DIRECTION. For example, by using

phrases such as “towards the pyramid”. Third, participants describe the MANNER of the route, also depicting MANNER and DIRECTION in gesture. Fourth, participants describe the DIRECTION of the route also depicting MANNER and DIRECTION in gesture. In these last two cases gesture is conveying information not present in speech. Therefore, while it seems as though gesture does mirror the information contained in speech, it regularly conveys additional information because of the conflation between way different semantic features are depicted in gesture.

These findings could be interpreted as providing evidence for both perspectives on gesture production. However, they reveal something interesting. The process of communicating with someone can be conceptualised as a coordination problem (Schelling, 1960; Clark, 1996) where interactants produce behaviours that reduce the amount of entropy in the world (Bara, 2010). This means that a communicator can either rely on novel, but transparent behaviours (such as gesture) or they can rely on common ground, including shared visual context by naming landmarks. Therefore, perhaps what these findings highlight is that participants either rely on the common ground afforded by the shared map features, or they rely on the transparent nature of gestural depictions.

In order to further explore these ideas, the rest of this section will focus on the relationship between gesture and higher level semantic features of participants construal of the route and maps.

#### 4.2.5 Frames of Reference

This section focuses on the invocation of frames of reference through speech. It is important to acknowledge that while a frame of reference is almost always present (the examples in chapter 3 attest to this), it can be explicitly invoked through language and/or gesture. For the map task, it is difficult to extract the frame of reference a particular individual was invoking through gesture because tracing gestures (as well as gestures that depict a position within gesture space) necessarily use gesture space as an encompassing secondary reference object, and therefore could be considered to employ a global frame of reference. However, tracing gestures are regularly produced from the producer’s perspective and thus could be described as employing a relative frame of reference. It is

for this reason that gestures have been tagged according to perspective adopted and not frame of reference (this will be covered in the next section). The frame of reference adopted in speech will be compared to the semantic categories employed in both speech and gesture. Additionally, the effect frame of reference has on the complexity of linguistic (vocal) performance will be explored.

Before pursuing the analysis of frame of reference it is worth exploring how frame of reference was tagged and providing some motivation for this tagging.

### Global

Semantic units were tagged as having a global frame of reference if they included an encompassing reference object. Typically, it is the world that people take as a secondary reference object, using cardinal directions to explain the relationship between objects (Levinson, 2003; Talmy, 2000a). In the map task, people did employ pseudo-cardinal directions, referring to top, right, bottom, and left of the map as north, east, south, and west respectively. For example:

- (4.1) G:           OK (.) so: out y- you do have the ecks start don't you  
       F:           yes I do  
       G:           so it comes out of that (.) pretty much west

In this case, “west” is being used to describe the direction that F is to draw the route, which, in egocentric terms, is towards the right. However, this is not the only way a global frame of reference is invoked. For example:

- (4.2) G:           and again the extreme of this kind of bend  
       F:           yeah  
       G:           is level with the ecks

In this second case, G is describing the “bend” as it goes around the lemon grove (which is directly underneath the start point). G describes the route at this point as being “level with the ecks”. The phrase “level with” invokes a global frame of reference because it

requires the conceptualisation of a line linking the start point and the route. It is this line which acts as an encompassing secondary reference object that is “level”, by which G means something like vertically beneath. This analysis follows from the idea that when people are waiting in a queue, it is possible to say something like “A is ahead of B (in the queue)”. This does not make sense unless the queue, which A and B must both form part of, is being used to determine the meaning of “ahead”, because “ahead” invokes a frame of reference in which the figures are encompassed by the queue (Talmy, 2000a, p. 204).

### **Intrinsic**

An intrinsic frame of reference was tagged whenever someone uses a specific part of a landmark as a reference object, but they are not also using an encompassing secondary reference object.

(4.3)

G:                   and then go down like (.) from the top of the limestone cliffs: (.) like  
                          let the line go down to where it says cliffs if you get what I mean

This example shows two instances of an intrinsic frame of reference. First, “from the top of” anchors the route to a particular point of a landmark. Words like “top”, “bottom” and “side” are one of the most common ways to invoke an intrinsic frame of reference. Another common strategy can be found in the second instance in this example. Here, G uses the placement of the words, which always appear under the landmark, to position the route. It is important to note that the particular part of the route G is describing does not start at the literal top of the limestone cliffs, but is on the right of the limestone cliffs by about two centimetres. Nor does the route actually touch the words underneath the landmark, but maintains the roughly two centimetre distance it had at the top. Therefore, it would be equally valid to describe the route as being “level with the top of limestone cliffs” or “in line with where it says cliffs”. For example, G’s turns in the following example were tagged with a global frame of reference:

(4.4)

- G: then (0.9) three centimetre- er: three or four centimetres (0.9) er::  
let's say three and a half centimetres south west of that beak
- F: okay
- G: and then three centimetres south of the word (.) of the letters can  
(0.7) cee ay ay (.) cee ay en

Here, G is using particular parts of the landmark “pelicans” as reference objects (i.e., the beak and letters within the word pelicans), but he is also using the map as a secondary reference object, through the use of cardinal directions. Therefore, in the terms of frame of reference, intrinsic can be embedded within global. The distinction between the last two examples highlights the fact that frame of reference is not an inherent feature of the object being described but is tied to the participants’ particular construal of that object (Langacker, 2013, p. 4).

### Relative

A relative frame of reference is invoked when a participant describes the route or another landmark using only an egocentric perspective. Usually, this is in terms of right and left but it is occasionally in terms of above and below (contrasted with top and bottom, which would be tagged as being intrinsic). For example:

(4.5)

- G: and you're gonna loop round that (0.8) so it's gonna have a nice left  
hand loop around it

Just prior to this turn, G had introduced the pelicans as the next landmark she is going to describe the route in relation to. Therefore, “that” and the final “it” are both references to the pelicans. Here there are two semantic units that could have been tagged for frame of reference. The first, “you're gonna loop round that”, does not include any explicit information that could be tagged as one of the frames or reference of interest

here. For this unit, F would have to work out the ORIENTATION of the route relative to the landmark. This was tagged as “none”. In the next unit, G actually elaborates on the orientation of the route to the landmark, stating “a nice left hand loop”. This unit would have been tagged as invoking a relative frame of reference since it does not involve a secondary reference object, nor does it pick out anything in the landmark that could help F work out the ORIENTATION of the route relative to the landmark. However, by left, G means her left and thus she also means F’s left as he looks at the map. F can work out the orientation of the loop because his egocentric perspective of his map is the same as G’s egocentric perspective of her map.

(4.6)

- F: inline with the ee of the avalanche got ya o:kay which is: p=fairly erm .hhh in=in my case (.) picket fence (.) just to the right of the ee of the final ee of picket fence
- G: right yeah
- F: which may be just to right of the final ee of broken gate
- G: it is just to the right <of ee of> broken gate yes
- F: ((imitates sound of gunfire)) done .hh right now we’re smooshing our way up in a generally northwest direction above lemon grove

There is a lot happening in this example, but it highlights two crucial points. F’s first turn includes the phrase “to the right of the ee of the final ee of picket fence”. This was tagged as invoking an intrinsic frame of reference since the mention of the “final e” allows G to work out the ORIENTATION of the route, even though this also adopts an egocentric perspective. Therefore, in the same way that intrinsic may be embedded within global, relative can be embedded within intrinsic. In his final turn in this example, F using the word “northwest” to describe the direction of the route and “above lemon grove” to describe its location. In the tagging scheme adopted here only “above” was tagged as invoking a frame of reference. Direction, unless it is direction relative to another landmark, is not tagged as invoking a frame of reference. Therefore, in this final



turn, the frame of reference adopted was tagged as relative, since “above” is relative to the perspective of the participant.

These examples highlight the key distinction between frame of reference that emerged as a result of using the scheme outlined in the previous chapter. The rest of this section explores the relationship between these terms and the semantic units found in the map task.

	<b>P</b>	<b>G</b>	<b>F</b>	<b>Dir</b>	<b>O</b>	<b>M</b>	<b>Dis</b>
<i>Intrinsic frame of reference (N = 385)</i>							
Speech	22(6%)	361(94%)	10(3%)	102(27%)	230(60%)	62(16%)	109(28%)
Gesture	40(10%)	9(2%)	2(<1%)	46(12%)	15(4%)	47(12%)	7(2%)
Both	1	9	0	22	11	19	6
<i>Global frame of reference (N = 444)</i>							
Speech	303(68%)	247(56%)	36(8%)	101(23%)	109(25%)	100(23%)	75(17%)
Gesture	66(15%)	13(3%)	1(<1%)	43(10%)	26(6%)	44(10%)	15(3%)
Both	49	8	0	20	16	24	10
<i>Relative frame of reference (N = 352)</i>							
Speech	18(5%)	242(69%)	20(6%)	110(31%)	279(79%)	60(17%)	51(15%)
Gesture	27(8%)	6(2%)	0(0%)	79(22%)	33(9%)	72(21%)	17(5%)
Both	7	4	0	45	28	38	8

Table 4.8: The distribution of speech and gesture by the different frames of reference

The first thing to notice is that there are generally lower incidences of gesture when frame of reference is taken into account, than when it is not. However, frame of reference is not a significant predictor of gesture ( $\chi^2(24) = 5.973, p = 0.1129$ ). Regardless, there are some interesting results at the level of semantic categories. First, there is a significant effect of frame of reference on GROUND in speech ( $\chi^2(15) = 21.112, p = < 0.001$ ). Paired contrasts between frame of reference types show that relative ( $\beta = 1.76 \pm 0.22(SE), z = 7.946, p = < 0.001$ ), intrinsic ( $\beta = 3.47 \pm 0.48(SE), z = 7.258, p = < 0.001$ ), and global ( $\beta = 1.02 \pm 0.23(SE), z = 4.521, p = < 0.001$ ) all had significantly higher incidences of GROUND in speech when compared to the conditions in which no frame of reference

was explicitly invoked. Within frames of reference, intrinsic had significantly higher incidences of GROUND in speech than both relative ( $\beta = 1.71 \pm 0.43, z = 3.944, p = < 0.001$ ) and global ( $\beta = 2.44 \pm 0.41, z = 5.984, p = < 0.001$ ). Finally, a relative frame of reference has significantly more GROUND in speech than a global one ( $\beta = 0.74 \pm 0.27, z = 2.704, p = 0.0345$ ). These findings are likely the result of the fact that intrinsic and global elements necessarily include a reference object that the route can be anchored to. This is also the reason behind the larger presence of POSITION in speech for the global frame of reference, which is determined by particular reference to the fixed structure of the map (regardless of perspective). This finding is also significant ( $\chi^2(15) = 36.092, p = < 0.001$ ). This time contrasts showed that a global frame of reference has a higher incidence of POSITION in speech than a relative one ( $\beta = 3.65 \pm 0.34(SE), z = 10.850, p = < 0.001$ ) and an intrinsic one ( $\beta = 4.01 \pm 0.56(SE), z = 7.324, p = < 0.001$ ). Global also has a significantly higher incidence of POSITION than when no frame of reference (none) is invoked ( $\beta = 3.63 \pm 0.55(SE), z = 7.910, p = < 0.001$ ).

Turning to the gesture, there is a significant difference in terms of DIRECTION ( $\chi^2(24) = 11.522, p = 0.009$ ). Paired contrasts reveal that only the contrast between global and none ( $\beta = -1.63 \pm 0.51, z = -3.231, p = 0.007$ ) and intrinsic compared to none ( $\beta = -1.12 \pm 0.38, z = -2.945, p = 0.017$ ) are significant. Importantly, a relative frame was not significantly different from none in terms of DIRECTION in gesture. In terms of MANNER, frame of reference was once again significant ( $\chi^2(15) = 18.224, p = < 0.001$ ). Paired contrasts showed that when compared to none, relative ( $\beta = -1.09 \pm 0.24, z = -4.626, p = < 0.001$ ), global ( $\beta = -1.51 \pm 0.29, z = -5.233, p = < 0.001$ ), and intrinsic ( $\beta = -1.19 \pm 0.28, z = -4.169, p = < 0.001$ ) all significantly reduced the amount of MANNER found in gesture. There were no significant contrasts between frames of reference.

These findings seem in line with the emerging findings. It is important to reiterate that what defines frame of reference is the presence of a reference object. As demonstrated, the presence of GROUND in speech is correlated with a reduction in gesture. This data shows an additional aspect to this story. While all frames of reference are related to

a significant reduction in MANNER in gesture, only a relative frame of reference was not tied to a reduction in DIRECTION in gesture. Why might this be the case? Intrinsic frames of reference involve the establishment of a particular feature of a landmark. This particular feature can be used to further anchor the route to the landmark reducing the need to disambiguate DIRECTION. In a global frame of reference, DIRECTION is non-ambiguous because the coordinate structure is absolute. A relative frame of reference is most ambiguous since it is based on the perspective of the producer. The question this raises is two-parted. First, why are DIRECTION gestures no less likely to occur with a relative frame of reference than when no frame of reference is invoked at all? And second, why are DIRECTION gestures less likely to occur when an intrinsic or global frame of reference is invoked? Answering the second first, it might seem logical that participants are using intrinsic or global frames of reference to position of the route or other landmarks, rather than describe it. However, if this were the case then it might be expected that there would also be less DIRECTION in speech associated with global or intrinsic frames of reference, and it turns out that DIRECTION in speech is not significantly affected by frame of reference ( $\chi^2(24) = 0.2436, p = 0.9703$ ). A more likely alternative is that people are not gesturing information regarding DIRECTION because, due to the shared visual context afforded by reference to the map, it is not necessary for them to do so. In this case, intrinsic and global frames of reference are related to the strategy outlined above which include GROUND and DIRECTION in speech but without the use of gesture. This is possible because DIRECTION can be gleaned from the speech. This strategic perspective on the inclusion of DIRECTION gestures is in line with a composite signal perspective on gesture. It also provides a possible answer to the first part of the question raised above. The reason for DIRECTION gestures with a relative frame of reference is that they are being used to communicate information about DIRECTION (albeit from a producer's perspective) because this is only ambiguously provided by speech. In other words, it is possible that an utterance producer can be surer that an interlocutor will understand their message when ambiguous features are also presented in gesture.

However, there is another key difference between intrinsic, global and relative frames of reference. The secondary reference object associated with a relative frame of reference

is the utterance producer's and/or comprehender's perspective. Therefore, it could be that the use of a relative frame of reference is tied to a heightened sense of embodiment.

#### 4.2.6 Perspective

Perspective is of crucial importance to the theory of gesture being developed in this thesis because the distinction between the different perspectives creates a natural contrast between gestures that are clearly for the producer and gestures that can be used during communication. In other words, perspective can be used as a natural version of mutual visibility experiments (Bavelas and Healing, 2013). What's more, there are different levels to which a comprehender is being invited to pay attention to gesture. The producer, in adopting an interlocutor's perspective, is explicitly inviting the comprehender to share the gesture's meaning. Therefore, we have two extremes: gestures that are not intentionally communicative (those produced with an unshared perspective) and gestures that are (those produced with a shared or second person perspective). The gestures that are produced from a shared perspective are arguably a case of intentionally shown behaviours (Wharton, 2009). In the middle are the everyday, standard gestures that are produced from a producer's perspective. The question then, is do first person gestures share features of gesture with an unshared or shared perspective?

Before exploring the data in the corpus, it is useful to explore the utilisation of perspective in an extended example.

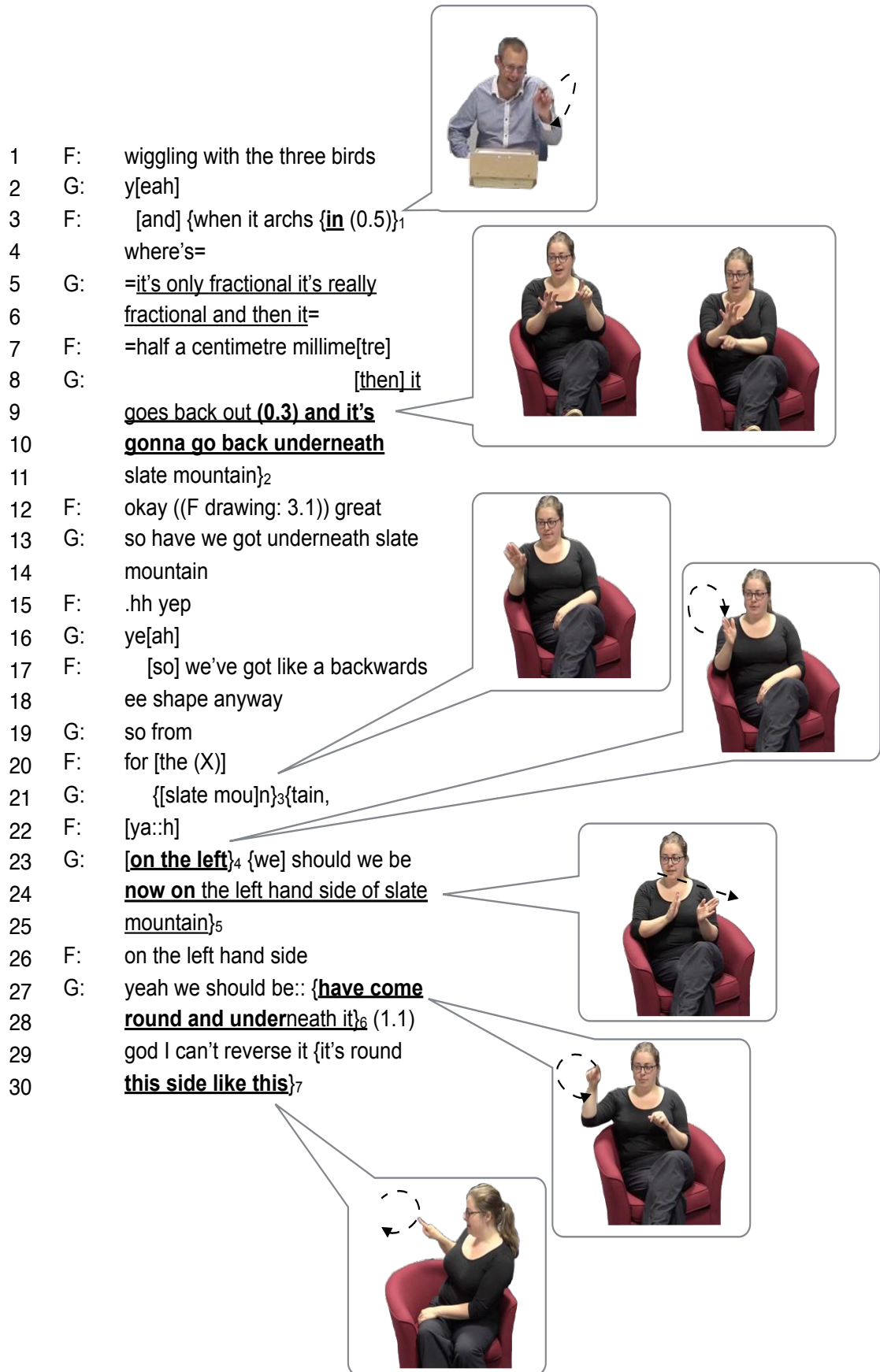


Figure 4.5: Example of different perspectives

In figure 4.5, there are several instances of different perspectives being utilised in the explanation of this section of the route. The first gesture appears on line 3 and accompanies F's align leading to his check on line 4 "where's". His gesture is from a first person perspective and its stroke is concurrent with the word "in". Also concurrent with the word "in", G places her hand in the arrangement seen in the left image of the second gesture. This arrangement is held in a pre-stroke hold from lines 3-9. In this two handed gesture, the right hand represents the slate mountain and the left hand depicts the DIRECTION the route is travelling. The stroke of this gesture, the endpoint of which is depicted in the image on the right, represents the route from F's perspective, and is therefore second person. The route has travelled from the right of the slate mountain to underneath the left of it. Following this, F begins to draw the section of the route. F bookends his drawing with "okay" and "great". On line 13-14, G produces an align, ensuring that F and G are at the same point. Following this is a two-part affirmation sequence followed by a F's check on lines 17-18. The next gesture appears in G's turn which is overlapped by F. G says "so from the slate mountain" producing a positioning gesture with her left hand (placed similarly to the left hand representing the slate mountain on lines 3-11). G's hand is placed prone with slightly elevated fingers. Her next gesture, which uses the same hand and is concurrent with "on the left" depicts a leftward movement (from G's perspective) whilst rotating her hand 90 degrees to a supine ORIENTATION with the thumb pointing upwards. This gesture is also produced from a first person perspective. G's next gesture on lines 23-25 involves both hands in a leftward movement. These last two gestures depict the ORIENTATION of the route relative to the slate mountain. F, who has not yet drawn the line going under the slate mountain, checks whether he has understood the instructions on line 26. G's answer begins on line 27 and is reformulated on lines 29-30. In her first attempt she says "we should be:: have come round and underneath it" the stroke of her gesture is concurrent with "have come round and under" and it depicts the circular path the route has travelled around the slate mountain. This gesture is realised in a first person perspective. Immediately following this gesture (in the 1.1 second pause), G looks up, as if to look at the trace she just produced. Upon

realising that her gesture was produced from her own perspective G explains that she is having difficulty producing gestures from a second person perspective and shifts to the shared perspective adopted in the final gesture. This gesture depicts the same route as the previous one, but is concurrent with a phrase that includes deictic reference to the trace “this side like this”. This gesture, therefore, would be labelled as speech framed.

Figure 4.5 shows how different perspectives can be realised over the course of the map task. It also highlights something important about producer’s awareness of their gestures. It is likely that G was not aware of her gesturing on lines 21-25, only becoming aware as a result of F’s problematisation of ORIENTATION. Even then she produced a gesture from her own perspective before switching to a shared one. Her motivation behind doing this requires some explanation. Earlier in the map task F was worried that by performing gestures from a shared perspective, G might somehow spoil the experiment. The result is that G began producing gestures from F’s perspective (all of the second person gestures presented below are from this task). However, it seems that the difficulty of maintaining the second person perspective was not worth anything gained by doing so, and G sticks to first person and shared perspective gestures from this point onward.

The rest of this section explores the deployment of perspective across the map task corpus.

	<b>P</b>	<b>G</b>	<b>F</b>	<b>Dir</b>	<b>O</b>	<b>M</b>	<b>Dis</b>
<i>Unshared Perspective (N = 432)</i>							
S	101(23%)	289(67%)	16(<1%)	125(29%)	152(35%)	107(25%)	56(13%)
<i>First Person Perspective (N = 611)</i>							
S	75(14%)	264(43%)	22(4%)	174(29%)	170(28%)	205(36%)	67(11%)
G	178(29%)	35(6%)	5(<1%)	335(55%)	76(14%)	348(57%)	62(10%)
B	53	22	2	149	52	169	29
<i>Second Person Perspective (N = 59)</i>							
S	8(14%)	20(34%)	2(3%)	23(39%)	6(10%)	20(34%)	7(12%)
G	16(27%)	9(15%)	2(3%)	33(56%)	9(15%)	41(70%)	2(3%)
B	6	3	1	15	3	18	2
<i>Shared Perspective (N = 118)</i>							
S	5(4%)	28(23%)	5(4%)	38(32%)	15(13%)	43(36%)	9(8%)
G	18(15%)	18(15%)	3(3%)	60(51%)	18(15%)	96(81%)	3(3%)
B	1	8	1	26	2	41	1
<i>No Perspective (N = 789)</i>							
S	188 (26%)	486 (62%)	45 (6%)	176 (23%)	294 (37%)	224 (29%)	177 (22%)

Table 4.9: Distribution of semantic categories by Perspective

Table 4.9 shows the distribution of semantic information relative to the different perspectives and demonstrates that there is a relationship between the presence of GROUND in speech and perspective. When unshared movements are produced, 67% of units contain GROUND information in speech. This is reduced to 43% with a first person perspective, 34% with a second person perspective, and 23% with a shared perspective. Further, when perspective is not considered then GROUND in speech is similar to unshared (62%). Statistically, perspective is a significant predictor of GROUND in speech ( $\chi^2(7) = 102.14, p = < 0.001$ ). Paired contrasts reveal that first person significantly reduces the incidence of GROUND when compared to none ( $\beta = -0.72 \pm 0.13(SE), z = -5.415, p = < 0.001$ ) and when compared to unshared ( $\beta = -1.05 \pm 0.14(SE), z = -7.453, p = < 0.001$ ). What’s more, unshared is not significantly different from none



( $\beta = -0.33 \pm 0.16(SE)$ ,  $z = -2.083$ ,  $p = 0.2276$ ). A shared perspective results in significantly fewer instances of GROUND in speech than first person ( $\beta = -0.87 \pm 0.25(SE)$ ,  $z = -3.428$ ,  $p = 0.0055$ ), unshared ( $\beta = -1.91 \pm 0.26(SE)$ ,  $z = -7.470$ ,  $p = < 0.001$ ) and none ( $\beta = -1.58 \pm 0.27(SE)$ ,  $z = -5.853$ ,  $p = < 0.001$ ). It is important to note that the lack of a finding for second person gestures is due to the fact low number of observations in this category. This points to a need for more research on perspective's affect on language and gesture.

These results follow the emerging trend in this chapter that gesture predicts a reduction in the incidence of GROUND in speech. Additionally, it suggests that unshared manual movements are similar to those that do not feature manual movements at all. Gestures with a first person perspective are, however, significantly different from unshared manual movements and when no gesture is produced at all. The suggestion then is that utterances produced with manual movements that are not shared behave, in relation to GROUND at least, like utterances that do not include gesture. Therefore, there is something distinct about the presentation of gesture in a place where it can be seen by the addressee.

It is also interesting to explore DIRECTION and MANNER in speech. It is clear from figure 4.9 that first person (36%), second person (34%), and shared perspectives (36%) can be grouped by MANNER in speech when compared to unshared (25%) and none (29%). However, the relationship between perspective and MANNER in speech is not significant ( $\chi^2(21) = 3.7756$ ,  $p = 0.4372$ ). Interestingly, DIRECTION is distributed in a different way, with first person having an identical distribution to unshared (29%), which is lower than second person (39%) and shared (32%) but higher than none (23%).

Turning to the gesture of the three perspectives that involve gesture (First Person, Second Perspective, and Shared). There is not a significant difference between them in terms of MANNER in gesture ( $\chi^2(10) = 0.8714$ ,  $p = 0.6468$ ) and DIRECTION is nearly identical across the different perspectives. However, table 4.9 suggests that while DIRECTION is evenly distributed across the different perspectives, there is more MANNER in gesture for second person and shared perspectives. Since DIRECTION and MANNER in gesture are often correlated, both being produced by tracing gestures simultaneously, this

suggests that MANNER is being gestured differently with a second person and shared perspective. Within the map task, it is common for gestures that do not include DIRECTION to be depicted using speech-framed gestures (McNeill, 2009) that include fixed gestures depicting the angle of the route, but do not depict the DIRECTION it is travelling in. The extract in figure 4.6 is an example of one of these types of gesture.

- 1 F: [yeah]  
 2 G: [per]fect .hhh {from there it's an  
 3 it's an=  
 4 F: and the ef of finish as well  
 5 G: it's another (0.6) **abrupt**<sub>1</sub>  
 6 {alteration in direction to be that  
 7 angle  
 8 F: right<sub>2</sub> this is eh mm heading  
 9 down towards ava=in the  
 10 direction of avalanche

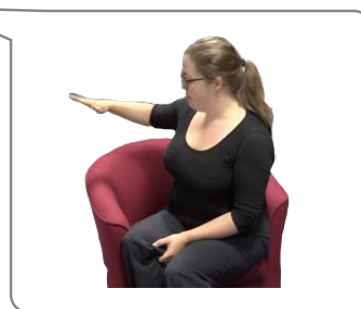


Figure 4.6: Shared Perspective without Direction

In figure 4.6, G is describing the route as it makes its way around pelicans. The route travels around the pelicans, starting above it and heading leftward and downward before turning and travelling back in the opposite direction. Once the route is directly beneath the centre of pelicans it begins a long incline rightwards and upwards towards the other side of the map. G has just finished explaining the leftward and downward section and is just about to begin explaining how it goes down and right to a point beneath and in the middle of the pelicans. Her instruction, on lines 2-7 begins by situating the instruction relative to what has already been explained (“from there it’s an”). F’s turn seemingly cuts off G’s, but continues the discussion of the previous topic, which related to the point the

route stopped at before it started heading right (which is in line with “f” of “finish”). G, who does not acknowledge F’s turn, describes the shift from heading rightwards to a leftward route as “abrupt”. Up until this point, beginning on line 2, G has been moving her arm with no discernable purpose. However, on the word “abrupt”, G jerks her hand upwards whilst holding her forearm away from her body. This gesture depicts the abrupt nature of the change in direction. Following this (on lines 6-8), G moves her arm into a position clear from her body and holds it for “direction to be that angle”. Here she draws attention to the ORIENTATION of her arm, producing a speech framed gesture. Her arm depicts the angle and therefore the MANNER of the route. However, these gestures are not necessarily tagged as depicting DIRECTION, because, from the gesture alone, it is impossible to know whether the route going upwards and leftwards or downwards and rightward. Perhaps coincidentally, F explicitly asks about the direction of the route, demonstrating that he is not sure of the exact direction the route is travelling in. Importantly, both the gestures are produced from a shared perspective. G has positioned her body so that her and F share a perspective on the ORIENTATION of her gestures.

If these gestures are more prevalent with a shared and/or second person perspective, then speech-framing should be correlated with perspective. This proved to be correct and perspective is a significant predictor of speech-framing ( $\chi^2(7) = 192.35, p = < 0.001$ ). Paired contrasts revealed that first person perspective included significantly fewer speech framed gestures than those gestures produced with a shared perspective ( $\beta = -2.01 \pm 0.36(SE), z = -5.639, p = < 0.001$ ).

These findings are important and will be the subject of future studies. However, the fact that there are more speech framed gestures produced with a shared perspective points to the fact that producers are clearly aware of what they are gesturing. This fact highlights the clearly intentionally nature of such gestures. Therefore, these results provide interesting insights into the intentional status of (more standard) first person gestures. It has been demonstrated several times above that speech with a high incidence of information pertaining to GROUND is correlated with a reduction in gesture. Here, we see that this is precisely the type of distribution that occurs with an unshared perspective. Speech adopting a first person perspective differed significantly in this respect and

it contained significantly fewer instances of GROUND in speech. If the hypothesis about the distribution of gesture and GROUND is correct then this would suggest that first person gestures are likely to be intentionally communicative. Moreover, this analysis goes against the view that gesture is produced with the primary purpose of helping the producer think about space. If that were the case there should be a distinction between those utterances without manual movements and those utterances with manual movements produced in an unshared environment. What's more there should be a distinction between the first person perspective and second person or shared in terms of gesture, and we should expect to find more similarities between first person and unshared. The fact these are lacking suggests that gestures produced with a first person perspective are communicative.

#### 4.2.7 Motion

Motion is a particularly interesting feature of the utterances because, as argued by Streeck (2009, p. 132) “[g]esture is motion and is therefore apt at depicting motion”. What's more, the motion inherent in an utterance is not always related to the motion of the referent. Like frame of reference, it is tied to the producer's construal of referent. This is exemplified in the distinction between the two following sentences:

- (1) The woman runs up the road to the house.
- (2) The path runs up the road to the house.

In sentence (1) it is the subject, in this case “the woman”, who is in motion and she is moving from an unspecified point on the road to a point further “up”—possibly the house. Couched in the terminology of Talmy (2000a), “the woman” who is acting as figure, is in a motion directed towards “the house”, which is acting as a primary reference object, but both are encompassed by another reference object, the road. However, in sentence (2) things are quite different. The path is not actually in motion, only a figure that traverses the path is in motion. In sentence two, the thing that might traverse the path is left unspecified and the motion associated with that figure is attributed to the path itself. In

this second case, the type of motion event described includes what is commonly referred to as “fictive” motion (Talmy, 2000a; Streeck, 2009; Langacker, 2013). As Streeck (2009, p. 136) puts it, fictive motion is where “[s]table features of the terrain are described in motion verbs that would describe what would happen to a person moving through it”.

The gestures found in the map task regularly trace the path as if it is moving within gesture space. The map task poses a situation where it is difficult to say whether or the motion described is fictive or not. This is because while the giver’s map has a static fixed route on it, the follower’s does not. The route is static but its replication is dynamic.

All semantic units in the map task were tagged for motion. Below are two examples that demonstrate how this was achieved.

**With Motion**

1 G: {okay and then you **go down**}<sub>1</sub>  
 2 {(0.4)[passed]}  
 3 F: [straight]  
 4 G: [yeah]<sub>2</sub> {go **go down** er: just  
 5 below}<sub>3</sub> {(0.5) where it says great  
 6 <viewpoint>}<sub>4</sub>

Figure 4.7: With Motion

The example shown in figure 4.7 includes two semantic units that were tagged as describing the route in motion. In this example, G is describing the very end of the route as it goes down the right of great viewpoint, before going under and then up the left hand side towards the finish. In this example, G describes the route as it goes down the right-hand side. On line 1, G produces a gesture which involves him bringing his left hand up with his index finger extended and bringing it down with the words “go down”. The gesture depicts the vertical nature of the route and is analogous with speech in that both relate to the direction of travel. This utterance is tagged for motion due to the use of the verb “go”. In the map task, it is difficult to judge whether or not this is a case of fictive motion since F will literally draw the route going down whereas the route of G’s map is static. Regardless, motion of the route is foregrounded in this utterance, whereas its fixed ORIENTATION to the rest of the map is ignored.

The end of G’s first turn (line 2) is in overlap as F questions the MANNER of the route with “straight”. In the pause that precedes the word “passed”, G starts to point directly at the map (probably touching the great viewpoint). This is an example of a gesture produced with an unshared perspective. He continues touching the map with his affirmative response to F’s check and begins a new gesture on line 4 with the word “go”. Following this, he repeats the word “go” so that the stroke of his gesture is once again concurrent with “go down”. This gesture is almost identical to the one he produced on line 1. Finally, he completes his turn by touching the map (not shown in figure 4.7) and saying “where it says great viewpoint”.

It is the use of the verb “go” that marks the semantic units in figure 4.7 as describing something in motion. Here, because the second person pronoun is used in the first instance, and implied by the imperative mood in the second, this is not taken as an example of fictive motion. It describes the motion of F as she draws her route.

Turning to non-motion and fictivity, in figure 4.8 there is an example which demonstrates several instances of non-motion semantic units.

1 G: if you can manage to get a slight  
 2 bend (.) a slight kink in it really it's  
 3 only slight (0.8) erm (.) between  
 4 lemon grove and pebbled shore  
 5 (.) {it seems like it just bends  
 6 **slightly** that way}<sub>1</sub> (.) it goes up  
 7 (.) a bit (.) it goes up steeper and  
 8 then it {**ends up**}<sub>2</sub>  
 9 (.) for the for the first half (0.9)  
 10 and then it bends and it's less  
 11 steep (.) in it's second half  
 12 between those two points (.)  
 13 between pelicans (.) and broken  
 14 gate (.) it's not directly straight  
 15 (0.6) it does have a bend in it

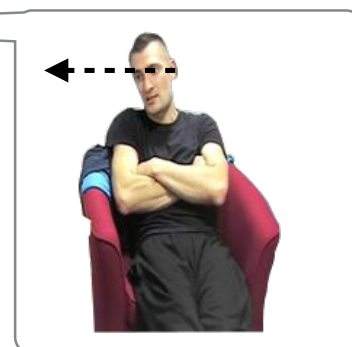
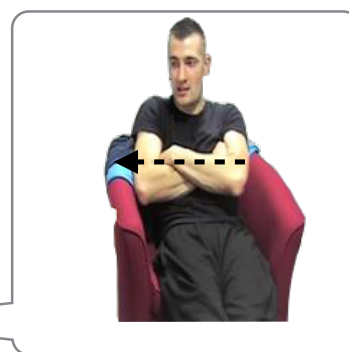


Figure 4.8: Without Motion

In figure 4.8, G is describing the route from beneath pelicans to above broken gate (which appears as picket fence on F's map). This section of the route represents the longest single trajectory on the route and has a slight upward bend in the middle of it. It is this bend that G is focussing on. Most of what G says here was not tagged as describing motion, because what he is describing is the bend on the line rather than the movement of a route as it emerges. For example, the noun phrases such as “a slight bend” (lines 1-2), “a slight kink” (line 2), and “a bend” (line 15), are static in the sense that they describe the shape of the route as if it is fixed. The clauses such as “it just bends” (line 5) and “it bends” (line 10) may at first seem to describe motion, but it is important to highlight that it is the route that should be in motion. Utterances including verbs like “bend” can only be applied to objects that have a fixed length and can be articulated (e.g., “he bent his arm” or “the branch is bent”). Other contenders for motion that are not tagged as motion are the intensive verbs used in this extract. For example, “it's less steep” (lines

10-11) and “it’s not directly straight” (line 14) once again describe the route as if it is fixed rather than actively describing its creation. However, there are two examples in figure 4.8 that were tagged for motion and those are the two instances of “it goes up” on lines 6 and 7. Here, “it” refers to the route and “goes” treats it as something that can move or is in the process of moving. Therefore, since the route is not actually moving for G, this is an example of fictive motion.

G does produce two gestures in this example, but only one of them was tagged as part of the corpus. In his first, he has his arms folded and his stroke occurs on the word “slightly” in the phrase “it just bends slightly that way”. Here, he directly references his gesture. The meaning of this utterance is that the bend in the line is rightward. His second gesture seems to follow the same path as his first, but this one involves his head. However this time it is produced with the words “ends up”.

These two examples highlight the way motion was tagged within the map task corpus.

	<b>P</b>	<b>G</b>	<b>F</b>	<b>Dir</b>	<b>O</b>	<b>M</b>	<b>Dis</b>
<i>Semantic units including motion (N = 775)</i>							
Speech	59(8%)	372(48%)	5(<1%)	429(55%)	218(28%)	305(39%)	66(9%)
Gesture	39(5%)	21(3%)	2(<1%)	323(42%)	28(4%)	336(43%)	18(2%)
Both	4	8	0	164	8	147	7
<i>Semantic units without motion (N = 1246)</i>							
Speech	329(26%)	721 (58%)	86(7%)	109(9%)	422(34%)	301(24%)	250(20%)
Gesture	181(15%)	42(3%)	8(<1%)	113(9%)	88(7%)	159(13%)	49(4%)
Both	58	25	4	30	50	87	25

Table 4.10: Distribution of semantic features with and without motion

The picture that is emerging in this analysis is that the presence of gesture is negatively correlated with the presence of GROUND information in speech. The presence of motion is positively correlated with the presence of gesture ( $\chi^2(8) = 12.912, p = <$



0.001) with motion being related to an increase in gesture ( $\beta = 1.0150 \pm 0.2283(SE)$ ,  $z = 4.447$ ,  $p = < 0.001$ ). Additionally, it is negatively correlated with the presence of `GROUND` in speech, however this relationship is not significant ( $\chi^2(9) = 3.3829$ ,  $p = 0.06587$ ). This suggests that gesture is tied to dynamic language. For the giver, the route is never in motion, it is a static object. For the follower, on the other hand, the route is produced through a drawing process. Tracing gestures depict static objects as if they are in motion. It seems that these gestures in the map task are likely to be non-static a depictions of the map and a route.

### 4.2.8 Complexity of Semantic features in speech

This section explores whether or not more complex speech is related to the presence of gesture and re-explores perspective in relation to complexity. Complexity is calculated by taking the number of semantic categories tagged in speech for each semantic unit. Importantly, only `POSITION`, `GROUND`, and `FIGURE` occur within each semantic unit more than once. This means that unlike previous data where the value associated with each semantic category within each semantic unit is binary, complexity represents the average incidence of the number of semantic categories tagged per semantic unit. The average across the whole corpus is 1.9 with a maximum of 7. This suggest that on average each semantic unit conveys information about two of the semantic categories explored in this analysis.

Therefore, if the emergent view about the position of gesture is correct and `GROUND` occurs without gesture because it is being used to anchor the route, then complexity should increase when gesture is absent. This is precisely what is found. Complexity significantly affects gesture ( $\chi^2(8) = 14.528$ ,  $p = < 0.001$ ) with complexity reducing the incidence of gesture ( $\beta = -0.35 \pm 0.08$ ,  $z = -4.473$ ,  $p = < 0.001$ ).

Combining complexity with the analysis from the previous section, it might be expected that there is a relationship between complexity and perspective. This turned out to be correct ( $\chi^2(8) = 71.103$ ,  $p = < 0.001$ ). Paired contrasts reveal that first person is associated with a significant reduction in complexity when compared to unshared ( $b = -0.39 \pm 0.06(SE)$ ,  $t(1543.69) = -6.334$ ,  $p = < 0.001$ ) and none ( $b =$

$-0.23 \pm 0.056(SE), t(1228.47) = -4.051, p = < 0.001$ ). What's more, there isn't a significant difference between unshared and none ( $b = -0.173 \pm 0.07, t(968.28) = -2.644, p = 0.0634$ ). This strengthens the idea that first person gestures are produced with significantly different speech. Ultimately, it suggests that first person gestures are being used communicatively.

### 4.2.9 Gesture's relationship to the word "around"

A prevalent early perspective on gesture focused on the use of gesture in relation to what was referred to as a lexical affiliate (Schegloff, 1984). The notion of lexical affiliate builds on the idea that a gesture is tied to the meaning of single lexical items and as a result is viewed in relation to that item. This idea has largely lost favour in gesture theory, because gesture and speech regularly convey information together that cannot be derived from either the speech nor the gesture alone.

A growing methodological thread is emerging that seems to have a similar underlying principle to that of lexical affiliate, particularly in the analysis of multimodal corpora. Generally, scholars working on corpora use searchable lexical strings to facilitate the exploration of those corpora, and this method has been adopted in the analysis of multimodal corpora. Two corpora have been analysed in this way. First, the Red Hen Lab Corpus (Lücking et al., 2013), which is a video corpus that contains 250,000+ hours of recorded communication, consisting of television broadcasts (including news interviews and advertisements) in a variety of languages. The Ren Hen Lab Corpus grows by another 150 hours everyday. This corpus, therefore, is one of the largest and prospectively most interesting corpora for gesture researchers. However, as with the analysis of any video data, there is an ever present question of how to access the information of interest (e.g., grammatical or semantic information). Typically with corpora of written texts this process is achieved through dedicated computer programs that tag tokens of words thus allowing analysis of collections of tokens. However, with video data this is not always possible and manual tagging is necessary. Because the Red Hen Lab corpus consists of television broadcasts, the video data comes with the subtitle track used during broadcast. However, since there is no subtitle track for gesture, the researcher can

only search the corpus using certain lexical strings (cf. Zima, 2014). This method was adopted by Zima (2014) who explored whether or not gesture was more prevalent with certain lexical strings. Zima argued that if gesture consistently co-occurs with certain lexical strings then gesture should be considered part of the grammar of that language. Another corpus that has been accessed using words is the Bielefield Speech and Gesture Alignment corpus (SaGA) (Lücking et al., 2013) which is a built corpus of multimodal interactions (e.g., video recorded direction giving in a virtual world). This corpus was built with gesture in mind and everything was coded manually. SaGA has also been used to suggest that gesture may be analysed as part of cognitive grammar (Kok and Cienki, 2016).

The crucial point is that these studies do not conceptualise spoken language as being primary to gesture, but use the tendency for speech to temporally and semantically align with gesture in order to explore the relationship between speech and gesture. The ultimate aim is to explore whether or not there are particular lexical items or constructions that co-occur with gesture (and in some cases those that do not). The purpose of such studies is to motivate the argument that gesture can be incorporated into a language's (cognitive) grammar.

An interesting pervasive finding from these studies is that the lexical item “(a)round” regularly occurs with gesture (Kok and Cienki, 2016; Zima, 2014). This subsection uses the lexical item “(a)round” as a methodological tool to explore the correlation between it and gesture. However, unlike those studies that have adopted such a methodology, this section will also explore the environment in which gesture does not occur. Before looking at the data as a whole it is worth exploring an individual example.

1 G: so it's still underneath slate  
 2 mountain, (.) **[but in line**  
 3 **with the –**  
 4 F: okay=  
 5 G: =**temple**<sub>1</sub> (0.6) er:m (.) [and  
 6 broken gate's sort of in the  
 7 **middle**]<sub>2</sub> (.)  
 8 [and then do a **loop round**  
 9 **again**]<sub>3</sub> (.)  
 10 [like **to the right and then**  
 11 **down**  
 12 F: hmm  
 13 G: **which is around where the**  
 14 **broken gate is**]<sub>4</sub>

Figure 4.9: Example of sequence including “around”

The example shown in figure 4.9 demonstrates how “around” is used during the map task. In G’s first utterance (lines 1-5), G is describing the location of the broken gate (an unshared landmark). She does so by first describing its position on the vertical axis as being “underneath slate mountain” and its position on the horizontal axis as being “in line with the temple”. Interestingly, the route is also underneath the temple and therefore F would have to enrich the meaning of “underneath” to something approximating IMMEDIATELY UNDERNEATH. During the first part of G’s first turn (lines 1-3) she is physically touching the page, however, concurrent with the word “but” she pulls her hand

backwards and downwards in a gestural depiction of “in line with”. At this point she had adopted a flat handshape in an oblique orientation. F’s backchanneled response comes in the middle of the spoken utterance which is cut off at “the” (perhaps indicating that G is thinking of the word “temple” since it appears as a pyramid on her map). However, it is important to note that F’s turn does not overlap with the gesture stroke.

G’s second turn (lines 5-11) can be broken into three components. Following G’s completion of her explanation started in her first turn, she begins to describe the route. This description can be broken into five semantic units based on the speech. She first describes the route ambiguously by stating (lines 6-7) “broken gate’s sort of in the middle” without explicitly saying what it is in the middle of. Here it is analysed as being in the middle of the, until now, unmentioned section of the route. In terms of semantic categories, the spoken component describes the position of the landmark relative to the route, which is acting as the **GROUND**. By inference therefore, this utterance describes the position of the route relative to the landmark. Concurrent with this description is a deictic gesture pointing to a particular point on G’s map, depicting the position of the broken gate not in gesture space, but on her map.

The second part of this description (lines 8-9) explicitly instructs F using the verb “do” followed by “a loop” acting as a noun phrase. In this analysis “round” is treated as a preposition post-modifying “do”, for which the noun phrase it pre-modifies is not realised explicitly. The enrichment required is that it is the broken gate that F needs to go around. Finally, “again” is a reference to the similar uses of this description that G has already made during the task rather than instructing F to make the same movement around the broken gate again, which would be impossible since this is the first time they have discussed the broken gate. The stroke of the gesture is concurrent with “loop round ag” highlighting the link between the gesture and route being described. The tracing gesture depicts a path which begins in the top left of gesture space, travels rightward gradually falling, and drops before coming back on itself slightly in a leftward direction. In terms of the semantic categories depicted in this gesture it clearly shows the **DIRECTION** and **MANNER** in which the line is travelling. As stated above, this analysis does not explicitly suggest that the gesture depicts **ORIENTATION** since there is

nothing in the gesture that depicts the GROUND around which the route is travelling. Instead the ORIENTATION must be inferred from the gesture and speech taken together as a composite signal.

The third, fourth and fifth semantic units in the speech are accompanied by a single gesture unit. The third (“to the right”) and fourth (“and then down”) semantic units describe the direction of the route, effectively breaking up the description contained in the second semantic unit. The gesture stroke is concurrent with third semantic unit and ends during the fourth (on the “d” of “down”). Importantly this gesture is almost identical to the one that co-occurred with the second semantic unit, suggesting that the third and fourth semantic units are an unpacking of the same idea. In the terms of McNeill and Duncan (2000), they are different unpacking of the same growth point. Further, here the gesture provides the same semantic information (i.e., DIRECTION and MANNER), but the speech provides direction information whereas before it provided MANNER information. The fifth semantic unit occurs with a post-stroke hold and reiterates what G has already said. However this time G makes explicit reference to “the broken gate” which is the landmark the route is going around. Therefore, here the speech describes MANNER of the route relative to a landmark as ground (importantly it is not analysed as describing the ORIENTATION since it cannot be known whether the route is going around the right or left from this utterance alone). The gesture, on the other, depicts the position of the final point of the route.

This analysis highlights several important characteristics of utterances including the word “around” that may be lost in corpora. First, “around” displays the property of principled polysemy (Tyler and Evans, 2003) and is underdeterminate by necessity (Atlas, 2005; Carston, 2002) rather than ambiguous. Following (Enfield, 2013; Enfield, 2009b), we can argue that “around” is a symbolic indexical requiring enrichment. This enrichment can either be provided by concurrent gesture or explicit spoken descriptions. We see both in the example depicted in 4.9. In the first use of “around”, the DIRECTION information is provided by the gesture that is concurrent with the lexical item. Second, the explicit spoken reference to the direction the route is travelling occurs in sequentially subsequent semantic units (the third and fourth) that do not themselves include a token

of the word “around”. Third, in the second realisation of the lexical token “around”, it is not accompanied by a gesture depicting DIRECTION but the hold phase of the gesture whose stroke co-occurs with the third and fourth semantic units. This time the semantic unit includes an explicit reference to the landmark that the route is going around. Although, this does not provide F with explicit information relating to the unspecified nature of around, it does help anchor the route to something that is in a known position on the map and therefore has already been established within their shared context. This analysis potentially highlights something important. Not only does it suggest that gestures depicting directional information co-occur with “around”, but that when that directional information is not supplied through gesture, producers provide it through other means. In this case, it is provided through explicit reference to a shared element. This leads to an alternative perspective to the view that gestures are grammatical. If “around” is lexically underspecified, then it is possible that producers adopt multiple strategies to direct their interlocutor’s enrichment of their utterance. In this case gesture is not grammatical but strategic, filling in the gaps left by symbolic-indexical items in the grammar. Following this example, there are potentially three different strategies for guiding enrichment:

1. Speakers guide the enrichment of “around” using gesture
2. Speakers guide the enrichment of “around” using explicit DIRECTION information
3. Speakers guide the enrichment of “around” by making reference to shared objects

The purpose of the rest of this section is to explore whether or not there is a distribution of strategies throughout the corpus. The proposed strategies lead to the following predictions:

1. When “around” occurs with gesture, producers will not explicitly mention a shared entity
2. When “around” occurs with explicit reference to a shared entity, producers will not gesture

3. Explicit mention to direction information in speech will not occur with the lexical token “around”

<i>Semantic units including the word “around” (N = 162)</i>							
	<b>P</b>	<b>G</b>	<b>F</b>	<b>Dir</b>	<b>O</b>	<b>M</b>	<b>Dis</b>
S	2(1%)	94(58%)	1(<1%)	30(19%)	34(21%)	155(96%)	10(6%)
G	7(4%)	2(1%)	0(0%)	74(46%)	5(3%)	74(46%)	0(0%)
B	0	0	0	16	2	73	0
<i>Semantic units not including the word “around” (N = 1859)</i>							
S	386(21%)	999(54%)	90(5%)	508(27%)	606(33%)	451(24%)	306(17%)
G	213(12%)	61(3%)	10(<1%)	362(20%)	101(6%)	421(23%)	67(4%)
B	62	33	4	178	56	161	32
<i>Semantic units including the word “around” but not including gesture (N = 82)</i>							
S	2 (2%)	61 (70%)	1 (1%)	13 (16%)	20 (24%)	79 (96%)	6 (7%)

Table 4.11: The distribution of semantic feature relative to the word “around”

Although gesture accompanies “around” 49% of the time, this relationship is not significant. Investigating the distribution of data that occurred with “around” reveals some interesting findings. First, gesture does not significantly affect GROUND in speech ( $\chi^2(8) = 193.17, p = 0.2446$ ). Nor does gesture significantly affect DIRECTION in speech ( $\chi^2(8) = 154.34, p = 0.3406$ ). This suggests that there is a different distribution of speech and gesture in the environment of “around”. As stated above the DIRECTION of the route as it occurs with “around” must be enriched. This enrichment will either be explicitly described in speech or provided by gesture. DIRECTION in speech is not significantly affected by the presence of “around” ( $\chi^2(8) = 0.7258, p = 0.3943$ ). However, “around” does affect the incidence of DIRECTION in gesture ( $\chi^2(8) = 1824.2, p = 0.01206$ ) with the presence of “around” increasing the incidence of DIRECTION in gesture ( $\beta = 1.1065 \pm 0.3661(SE), z = 3.023, p = 0.0025$ ).

These findings suggest that “around” is being significantly enriched by DIRECTION information in gesture, but not in speech. These findings do not rule out the fact that



gesture accompanying the word “around” should be considered part of the grammar of English. However, nor does it rule out the fact that gesture is just one way of enriching the meaning of “around”. More data might reveal the relationship between “GROUND” in speech and gesture in the environment of “around”. However, what this does suggest is that the gesture is being used to enrich the word “around”.

### 4.3 Interaction

Up to this point, the data has only been analysed in terms of how it contributes to signalling. According to the Clarkian action ladder, at level three signals are produced and recognised before level four at which projects are proposed and considered. Here, proposal and consideration are explored in relation to how they are tied to gesture. The move coding adopted in this task is not related to the signal but to the proposal, therefore by exploring the interrelation between move type and gesture, it is possible to glean information regarding the different ways in which gesture is being used during the act of proposing. In terms of considering, it was described in the chapter 3 how the map tasks were tagged for whether a move results in a secondary move that closes a project (e.g., ACKNOWLEDGEMENT or one that embeds a new project within an old. Therefore, if a move is of the closing sort, this suggests that a project has been finished and both participants are satisfied sufficient for current purposes (Clark, 1996). However, if a move is of the embedding sort, then some elaboration, either requested or not, has been provided. Theoretically, moves can be embedded *ad infinitum*, and each embedded creates a new level, however in the map task the highest level (or lowest, depending on one’s perspective) reached was 11.

This section explores the use of gesture in relation to Moves and Level.

#### 4.3.1 Moves and gesture

It is important to highlight that the following results are not the results for the number of moves in the map task. Instead, they represent the moves containing semantic units coded as representing one or more of the semantic categories outlined above. Each semantic unit is coded as belonging to a move and therefore a single move is regularly

broken into multiple units. Further, this section only includes those moves that are coded as containing semantic units<sup>5</sup>. That being said, a potential confound of the hypothesis suggested in the previous section is related to whether or not the particular move a participant is producing is responsible for the apparently strategic deployment of gesture.

This section explores the deployment of gesture, and more particularly, the distribution of gesture which has emerged above, and its relation to the different types of move. As demonstrated above, the type of move used does have an effect on the amount of gesture produced. The picture emerging in this chapter is that **GROUND** in speech is negatively correlated with **DIRECTION** and **MANNER** in gesture. If this is correct, then it can be expected that this relationship will be significant regardless of move type.

This section explores two questions.

- Which move types are more likely to occur with gesture?
- Is **GROUND** in speech negatively correlated with **MANNER** and direction in gesture regardless of move type?

Before attempting to answer these two questions, it is useful to explore the distribution of move types within the corpus generally.

<b>Move</b>	<b>Number (%)</b>
Instruction	670 (33%)
Check	406 (20%)
Clarify	323 (16%)
Align	178 (8%)
Explain	162 (8%)
Reply-Wh	105 (5%)
Acknowledge	64 (3%)
Question-Yes/No	50 (2%)
Reply-Yes	29 (1%)
Question-Wh	21 (1%)

---

<sup>5</sup>This would exclude most responses to yes/no questions, for example.

Ready	11 (1%)
Reply-No	1 (<1%)
Uncategorised	1 (<1%)

Table 4.12: breakdown of move types

Table 4.12 shows the distribution of move types across the corpus. From this table it can be gleaned that INSTRUCTIONS, CHECKS and CLARIFIES make up 69% of the corpus. This suggests that the majority of time spent communicating involves the production of INSTRUCTIONS, the questioning of INSTRUCTIONS and the CLARIFICATION of INSTRUCTIONS in response to CHECKS. As was demonstrated above, move type has a significant affect on the amount of gesture ( $\chi^2(13) = 29.317, p = 0.001107$ )<sup>6</sup>. The estimated intercepts for each move type reveal that ACKNOWLEDGEMENTS, CHECKS, EXPLANATIONS, and READY moves all result in a reduction in gesture. All other moves are associated with an increase in gesture. However, only CHECKS were associated with a significant decrease based on parameter estimates ( $\beta = -0.73 \pm 0.3(SE), z = -2.130, p = 0.0332$ ). Interestingly, REPLY-YES was associated with a significant increase in gesture ( $\beta = 1.36 \pm 0.6, z = 2.292, p = 0.02$ ). While the positive increase in gesture with yes responses might seem odd, it is important to reiterate that the data represented in this analysis, only includes those instances where space was being depicted or described. In relation to YES REPLIES, they were tagged on two occasions. First, they were tagged when a producer uttered the word “yes”, and second they were tagged when a producer uttered the word “yes” and/or repeated, verbatim, what had just been said. It seems that there were occasions where gestures were being produced with both. However, this does not mean that gestures regularly occur with yes responses only that those responses that were tagged for containing semantic information were more often than not gestural. Paired contrasts were used to break down the results and demonstrated that the only significant contrasts involve CHECKS and INSTRUCTIONS. CHECKS result in significantly fewer gestures than ALIGNS ( $\beta = -1.3 \pm 1.3, z = -5.303, p < 0.001$ ), CLARIFIES ( $\beta = 1.09, \pm 0.21, z = -5.282, p = < 0.001$ ), and INSTRUCTIONS ( $\beta = -1.271 \pm 0.19, z = -6.6, p = < 0.001$ ). Also, there are significantly fewer gestures with EXPLANATIONS than

---

<sup>6</sup>Only those moves with more than one occurrence form part of this analysis.

INSTRUCTIONS ( $\beta = -0.76 \pm 0.23, z = -3.242, p = 0.0466$ ).

These results show that few contrasts actually reveal a significant difference in terms of move type. However, these results do point to an interesting finding. CHECK moves are associated with a reduction in gesture, particularly when compared to CLARIFICATIONS and INSTRUCTIONS. CLARIFICATIONS and INSTRUCTIONS represent the primary move type through which task relevant information is given. Crucially, these move types are associated with a transfer of information, typically from the giver to the follower. CHECKS, unlike WH-QUESTIONS, involve a participant asking about already presented information. In other words the significant results relate directly to the giving a receiving of new information that is crucial to task completion. This suggests that gestures are more likely to be used when new information is being provided than when it is being requested. However, this does not explain the difference between INSTRUCTIONS and EXPLANATIONS. As described in the chapter 3, EXPLANATIONS are associated with descriptions of features of the map that do not constitute INSTRUCTIONS. In other words, EXPLANATIONS involve the description of landmarks and distances between landmarks. It has been demonstrated several times that utterances containing GROUND in speech are less likely to occur with gesture. Therefore, this finding makes sense and builds on the idea that it is the intentional transfer of (new) information as part of the task that leads to the use of gesture.

This result, however, is based on the wholesale presence or absence of gesture. What has been shown up to this point is that MANNER and DIRECTION are critical aspects of gesture and that they seem to negatively correlated with the presence of GROUND in speech. This section will further explore whether or not move type has a significant effect at this more specific level. Table 4.13 shows the distribution of semantic categories relative to move type<sup>7</sup>.

---

<sup>7</sup>Only those move types related to higher than 100 tokens have been included.

	<b>P</b>	<b>G</b>	<b>F</b>	<b>Dir</b>	<b>O</b>	<b>M</b>	<b>Dis</b>
	INSTRUCTION						
S	107 (16%)	360 (54%)	20 (3%)	268 (40%)	229 (34%)	245 (37%)	105 (16%)
G	50 (8%)	10 (2%)	3 (<1%)	196 (29%)	27 (4%)	221(33%)	28(4%)
B	16	2	1	194	13	111	14
	CHECK						
S	61 (15%)	219 (54%)	7 (2%)	96 (24%)	136 (34%)	133 (33%)	66 (16%)
G	23 (5%)	12 (3%)	1 (<1%)	64 (16%)	18 (4%)	69 (17%)	2 (<1%)
B	8	7	0	22	10	40	2
	CLARIFY						
S	55 (17%)	164 (51%)	8 (2%)	81 (25%)	86 (27%)	125 (39%)	72 (22%)
G	36 (11%)	17 (5%)	1 (<1%)	71 (22%)	25 (7%)	90 (28%)	14 (4%)
B	10	9	0	31	10	48	8
	ALIGN						
S	22(12%)	119(69%)	5 (3%)	30(17%)	65(37%)	43(24%)	19(11%)
G	44(25%)	7(4%)	0 (0%)	27(15%)	10(6%)	32(18%)	7(4%)
B	5	5	0	9	7	9	2
	EXPLAIN						
S	74(46%)	78(48%)	30(19%)	18(11%)	44(27%)	21(13%)	24(15%)
G	24(15%)	5(3%)	0(0%)	23(14%)	13(8%)	23(14%)	9(6%)
B	12	2	0	11	6	11	4
	REPLY-WH						
S	26(25%)	63(60%)	8(8%)	22(21%)	41(39%)	13(12%)	13(12%)
G	16(15%)	6(6%)	5(5%)	15(14%)	14(13%)	16(15%)	3(3%)
B	6	3	3	11	6	4	2

Table 4.13: Distribution of semantic information by Move type

First off, does move type have an effect on GROUND in speech? There is a significant effect of move type on GROUND in speech ( $\chi^2(13) = < 0.001$ ). Almost all move

types were associated with an increase in GROUND information in speech. The only non-significant result is from EXPLANATIONS ( $\beta = 0.57, \pm 0.33, z = 1.7, p = 0.08$ ). A potential reason behind this result is that EXPLANATIONS are typically used to describe the location of something by describing its position. Looking at 4.13, this notion is corroborated by the high incidence of position (46%) associated with EXPLAIN moves. Investigating contrasts between move types reveals no significant differences. These results suggest that, on the whole, there was no significant variation in terms of the amount of GROUND information in speech based on move type.

What about MANNER information conveyed in gesture? Once again there was a significant effect of move type on MANNER in gesture ( $\chi^2(13) = 96.216, p = < 0.001$ ). Unlike GROUND in speech this was tied to particular move types and not all resulted in an increase of MANNER information conveyed through gesture. CHECKS ( $\beta = -0.9 \pm 0.4, z = -2.323, p = 0.02$ ) and EXPLAINS ( $\beta = -1.2 \pm 0.5, z = -2.693, p = 0.007$ ) are both significantly tied to a significant reduction in MANNER in gesture. Contrasts revealed that instruction moves had significantly more semantic units contain MANNER in gesture than CHECKS ( $\beta = 1.5 \pm 0.2, z = -6.211, p = < 0.001$ ) and EXPLANATIONS ( $\beta = 1.76 \pm 0.31, z = -5.539, p = < 0.001$ ).

Turning to DIRECTION in gesture, there was a main effect of move type ( $\chi^2(13) = 66.441, p = < 0.001$ ). However, looking more closely at the estimated parameters does not reveal any significant effects of moves other than ACKNOWLEDGEMENT and YES REPLIES. INSTRUCTION moves are almost significant ( $\beta = 0.84 \pm 0.43, z = 1.918, p = 0.055$ ). This suggests that DIRECTION gestures are more likely to occur with INSTRUCTIONS than any other move. Contrasts reveal that, once again, INSTRUCTION moves had significantly more semantic units contain DIRECTION in gesture than CHECKS ( $\beta = 1.1 \pm 0.2, z = -4.756, p = < 0.001$ ) and EXPLANATIONS ( $\beta = 1.45 \pm 0.32, z = -4.550, p = < 0.001$ ).

These findings suggest that not only is gesture more prevalent with instruction moves than checks and explanations, but that it is also the case when direction and MANNER gestures are considered.

Although these results are interesting and point to the possibility that different move types are tied to the presentation of gesture, it also present a potential confound for the

results described earlier. Up until this point, it has been suggested that DIRECTION and MANNER in gesture occur in different scenarios to GROUND in speech. Although the move types described here are not tied to any significant contrasts in terms of GROUND in speech, this was not the case for DIRECTION and MANNER in gesture. It is possible therefore, that what the findings have highlighted so far is the relationship between checks and explanations on the one hand and instructions and clarifications on the other. In order to investigate this possibility the rest of this section will be spent exploring the data associated with explanations, instructions, checks, and clarifications.

### Explanations

The question is whether or not MANNER and DIRECTION in gesture are negatively correlated with GROUND in speech in the environment of explanations. GROUND in speech is predicted by MANNER in gesture in the environment of explanations ( $\chi^2(5) = 14.294, p = < 0.001$ ). Parameter estimates reveal that MANNER in gesture is responsible for a significant reduction in GROUND in speech ( $\beta = -3.3324 \pm 1.0983, z = -3.034, p = 0.00241$ ). Turning to DIRECTION in gesture, this also predicts the presence of GROUND in speech ( $\chi^2(3) = 25.729, p = < 0.001$ ). Estimated parameters reveal that DIRECTION in gesture leads to a reduction in GROUND in speech ( $\beta = -3.5015 \pm 1.0780, z = -3.248, p = 0.00116$ ). Therefore, these results show that the distribution of GROUND in speech and MANNER and DIRECTION in gesture is the same in the environment of explanations as the general trend in the entire corpus.

### Instructions

Within the environment of instructions GROUND in speech is predicted by MANNER in gesture ( $\chi^2(3) = 56.977, p = < 0.001$ ) with MANNER in gesture being responsible for a reduction in GROUND in speech ( $\beta = -1.4849 \pm 0.2, z = -7.416, p = < 0.001$ ). Ground in speech is also predicted by DIRECTION in gesture ( $\chi^2(5) = 8.0766, p = 0.004484$ ) with DIRECTION in gesture predicting a reduction of GROUND in speech ( $\beta = -1.0353 \pm 0.2602, z = -3.979, p = < 0.001$ ).

**Check**

It is the same case for checks. GROUND in speech is predicted by MANNER in gesture ( $\chi^2(5) = 11.88, p = < 0.001$ ) with MANNER in gesture reducing GROUND in speech ( $\beta = -1.4321 \pm 0.2958, z = -4.842, p = < 0.001$ ). This is also the case for DIRECTION in gesture ( $\chi^2(5) = 8.0882, p = < 0.004455$ ) with DIRECTION in gesture reducing the presence of GROUND in speech ( $\beta = -1.0984 \pm 0.3198, z = -3.434, p = < 0.001$ ).

**Clarify**

Last, the same situation holds for clarifies. MANNER in gesture has an effect on GROUND in speech ( $\chi^2(5) = 8.0734, p = 0.004492$ ) with MANNER in gesture reducing the incidence of GROUND in speech ( $\beta = -1.1329 \pm 0.2636, z = -4.297, p = < 0.001$ ). DIRECTION in gesture behaves in the same way ( $\chi^2(3) = 6.3466, p = 0.01176$ ) with DIRECTION in gesture reducing GROUND in speech ( $\beta = -0.6946 \pm 0.2745, z = -2.531, p = 0.0114$ ).

These results demonstrate that while move type does have an effect on the incidence of gesture, within those most responsible for the change in gesture the same relationship found in the rest of the map task still holds.

**4.3.2 Sequential position of gesture**

The map task data was coded for the level of a move as it occurs during the interactive sequence. This subsection explores whether or not gesture is more likely to occur at deeper levels in the interaction.

The graphs shown in figure 4.10 depict the effect of level on gesture, GROUND in speech, and MANNER and DIRECTION in gesture.

Analysing the effect of level on gesture showed that level does have a significant effect ( $\chi^2(11) = 20.894, p = 0.01313$ ). Estimated parameters reveal that level 3 ( $\beta = -0.2782 \pm 0.1212, z = -2.294, p = 0.02178$ ), level 9 ( $\beta = -0.8011 \pm 0.2734, z = -2.931, p = 0.00338$ ) and level 11 ( $\beta = -0.8011 \pm 0.3751, z = -2.931, p = 0.03268$ ) all result in a reduction in the presence of gesture. This suggests that the observation from the data that initially, as moves are embedded people gesture less. However, the apparent increase in gesture after level 13 is not realised statistically.



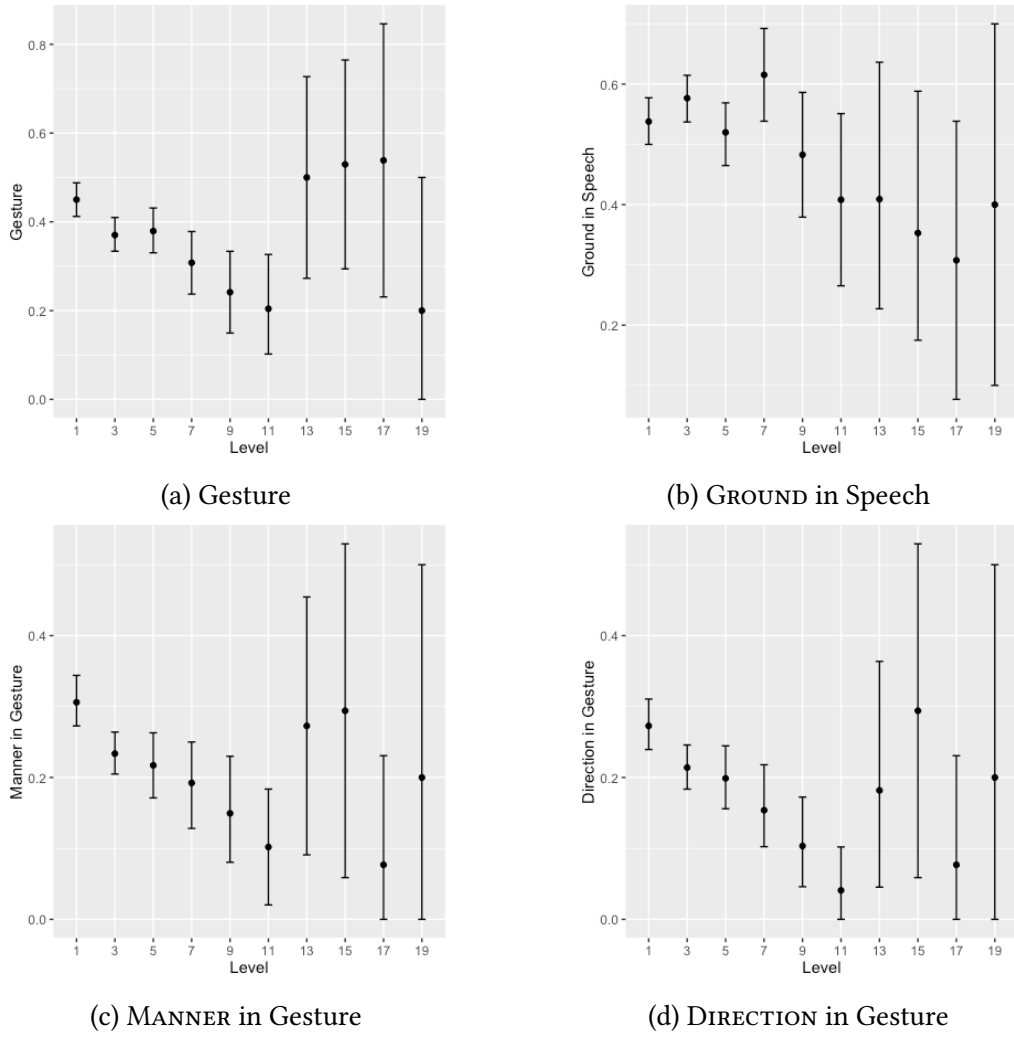


Figure 4.10: Plots of Level

Turning to MANNER in gesture, the figure seems to show that MANNER in gesture is similar to gesture generally. This is borne out statistically ( $\chi^2(11) = 17.189, p = 0.04584$ ). This time, level 3 ( $\beta = -0.28364 \pm 0.13215, z = -2.146, p = 0.03185$ ), level 5 ( $\beta = -0.34537 \pm 0.16839, z = -2.051, p = 0.04026$ ), level 9 ( $\beta = -0.77746 \pm 0.32338, z = -2.404, 0.01621$ ) and level 11 ( $\beta = -0.96946 \pm 0.49853, z = -1.964, p = 0.04957$ ) are all related to a reduction in the amount of MANNER in gesture.

Once again, DIRECTION in gesture behaves in the same way ( $\chi^2(11) = 18.413, p = 0.03067$ ). This time, however, only level 9 ( $\beta = -0.9059 \pm 0.3714, z = -2.439, p = 0.01472$ ) and level 11 ( $\beta = -1.6809 \pm 0.7328, z = -2.294, p = 0.02180$ ) are related to a significant reduction in DIRECTION in gesture.

Finally, GROUND in speech seems to be in a very different distribution relative to level. However, it does seem that there is an effect of level on GROUND in speech ( $\chi^2(11) = 18.626, p = 0.02857$ ). However, this time only level 11 ( $\beta = -0.6629 \pm 0.3077, z = -2.155, p = 0.0312$ ) is related to a decrease in GROUND in speech. However, unlike with the other analyses level 3 and level 7 are actually related to an increase in GROUND in speech.

These data are difficult to interpret, and should be interpreted cautiously. However, what they seem to show is that gesture (including MANNER and DIRECTION) are reduced as moves are embedded. With GROUND in speech it is not all clear what is happening as the level gets higher.

Therefore, these results potentially reveal something interesting. People gesture less when they elaborate on instructions that have already been given. This finding is in need of future research. However, it leads to a question (which cannot be answered with this data). Is the reduction in gesture a result of the fact that moves with gesture are less likely to require elaboration or that participants use less gesture when they are elaborating on any more (regardless of whether it included gesture)?

The first of these questions relates to the communicative import of gesture (i.e., how well gestures communicate) and the second question relates to the intentionally communicative nature of gesture. If it turns out that producers gesture less when they are elaborating or clarifying, then this might suggest that as people think more about the

content of their utterances, they gesture less. In other words, when people attempt to clarify their instructions they gesture less.

## 4.4 Summary

To summarise the data analysed in this chapter, there are many interesting findings that seem to highlight the distribution of gesture and speech, and, more specifically, the semantic features in the two modalities. Many of these findings are open to interpretation and can be used as evidence for either the theory that gestures are for the producer or that gestures are for the comprehender.

In terms of semantic features the most prevalent feature was that there seemed to be at least two environments found in utterances for the map task. First, people described **GROUND** by referring to distinct landmarks. In this environment participants were no less likely to refer to the **DIRECTION** of the route (through speech). This suggests that participants use the landmarks on the map to anchor the route. However, the environment in which people gesture seems to be in a negatively correlated to the environment in which people describe **GROUND**. **DIRECTION** in speech is correlated with **DIRECTION** and **MANNER** in gesture, but **DIRECTION** in speech does not occur with **MANNER** in speech. This suggests that the description of **DIRECTION** in speech is enriched by **MANNER** in gesture. In these cases, it seems that rather than anchor the route to landmarks, participants are anchoring the route to gesture space. Participants therefore, seem to adopt two strategies for describing the map. They either use objects in common ground, invoking a visual environment that is indirectly shared. It is indirectly shared because they are looking at two versions of the same map. The second strategy is to use gesture space as a representation of map. The gesture also enriches the speech. This fact is further evidenced by the use of depictions of **DIRECTION** in gesture. This distribution suggests that gesture is being used communicatively since it is not simply the case that semantic information in gesture is realised as a result of that information simultaneously being realised in speech. It is tied to a lack of crucial communicative information (i.e., common ground) not being explicitly referred to.

Turning to frames of reference, **DIRECTION** information in gesture does not occur

with utterances that invoke an intrinsic or global frame of reference. This is not the case with a relative frame of reference. There are at least potential reasons why this might be the case. First, it could be that the egocentric perspective invoked by a relative frame of reference heightens the embodiment associated with utterance. If this is the case then directional gesture could be a natural accompaniment of such utterances. The alternative perspective is that the gesture is being used to disambiguate the non-specific directional nature of the utterance. Compared to intrinsic and global frames of reference, which do not require perspective taking, a relative frame of reference involves the comprehender adopting the perspective of their interlocutor. It is possible that gestures accompanying relative frames of reference are used to facilitate this process.

Perspective represented one of the strongest pieces of evidence that gestures are being used communicatively. It was shown that manual movements produced in a unshared environment (i.e., directly on the map) accompany language that is not qualitatively different from utterances that do not include any form of manual movement. Gestures produced with a first person perspective accompany language that is qualitatively different from both utterances that do not include any manual behaviours and those that include manual movements produced from an unshared perspective. What is more, gestures produced from a first person perspective are significantly less complex than unshared manual movements and utterance that do not include manual movements. This suggests that it is shared manual movements (i.e., gestures) and not manual movements that have an effect on the concurrent spoken language.

Turning to the interactive properties of gesture, it seems as though gesture is more likely to occur with moves that represent information over those that request it. Furthermore, gestures are more likely to occur with crucial task relevant contributions. Finally, the more embedded a project is the less likely it is to occur with gesture. This is hard to interpret, but it is possible that as people try to clarify or elaborate on their utterances they gesture less because they are more actively thinking about the content of what they are saying and inhibiting gesture. However, it is also possible that those turns that are accompanied by gesture are less likely to result in checks or elaboration. The one thing that does seem evident is that people are not producing more gesture in embedded

projects.

Therefore, the findings of this chapter largely point to the idea that gestures are produced communicatively. These findings and the findings of the next two chapters will be picked up in the discussion chapter below.



## **Part III**

# **Gesture and Comprehension**





# Chapter 5

## Methodology 2: The Visual World

### Paradigm

#### 5.1 Introduction

The previous chapters have highlighted the fact that people describing space frequently employ gesture in a meaningful way, contributing to the overall message that is being conveyed. This chapter and the next focus on the effect of gesture on the real-time comprehension of utterances. In order to answer this question it is necessary to explore what effect gesture has at the local semantic level or level three (signalling and comprehension) of the Clarkian action ladder (see figure 5.1).

Level	Utterer A's actions	Addressee B's actions
4	A is <i>proposing</i> joint project $w$ to B	B is <i>considering</i> A's proposal of $w$
3	A is <i>signaling</i> that $p$ for B	B is <i>recognizing</i> that $p$ from A
2	A is <i>presenting</i> signal $s$ to B	B is <i>identifying</i> signal $s$ from A
1	A is <i>executing</i> behavior $t$ for B	B is <i>attending</i> to behavior $t$ from A

Table 5.1: Action ladder involved in language use (Clark, 1996)

It is also important to recognise that at a semantic level gesture is not an all or nothing affair. People employing gesture tend to use it in a way that complements what they are expressing through speech rather than expressing identical content. In this way the semantic content associated with the object being described is often distributed across

speech and gesture (Beattie and Shovelton, 2006). However, it is not thought to be the case that speech and gesture are derived from an identical store of representations, with gesture relating to what is not described in speech and speech relating to what is not described though gesture—a view sometimes referred to as the firehose theory (McNeill, 2012, p. 187). McNeill attributes the firehose theory to (de Ruiter, Bangerter, and Dings, 2012a; De Ruiter, 2007). The firehose theory is identical to the tradeoff hypothesis and thus states that as speech becomes more difficult people gesture more and as gesture becomes difficult people speak more. The term ‘firehose’ is used because, due to water pressure a firehose crimped at one place will cause a bulge at another. Similarly, communication, when crimped in terms of speech will form a bulge through a different communicative modality, which in this case might be gesture. McNeill’s view, however, is that speech and gesture are co-expressive in that they are related to the same idea unit, which is itself multifaceted (McNeill, 2012). This means that while gesture is communicative, one should expect redundancy. With this in mind it is important to explore how a single semantic element is comprehended when it is expressed through both speech and gesture, speech only, gesture alone, and neither speech nor gesture. From a pragmatic perspective it is important to consider the reasons why and when addressees pay attention to gesture. Within the gesture literature, it has been shown that gesture is integrated very early on in the comprehension process (Kelly, Özyürek, and Maris, 2010), and that gesture presents a range of benefits for the comprehender. The standard idea in pragmatics is that people pay attention to a communicative behaviour because it is intended to communicate. This idea has been captured in the post-Gricean distinction between communicative and informative intentions whereby an individual who desires to communicate some information will direct their interlocutor to this information by producing some behaviour that makes that information manifest or salient (Sperber and Wilson, 1995). Placing an empty glass in someone’s field of vision can be used to communicate that the person to whom the glass belongs would like it filled up in the same way that an utterance of “I’m really thirsty” can—even though neither of these behaviours explicitly expresses that information. The reason why people pay attention to communicative behaviours is standardly explained through one of two principles. Ac-

According to Gricean and neo-Gricean scholars the co-operative principle posits that participants search for information in behaviour because interlocutors are believed to be co-operative, in other words, although on the surface it may not initially seem like the behaviour has a directly decodable meaning, processing it will lead to the comprehender inferring information the utterer intended to communicate (Levinson, 1983; Clark, 1996). The alternative, posited by Relevance Theorists is the Communicative Principle of Relevance which states that communicative behaviours come with a presumption of their own optimal relevance, which means that humans come equipped with a predisposition to pay attention to communicative behaviour (Sperber and Wilson, 1995). This captures the fact that humans find it impossible not to treat speech sounds as communicative. These points raise the question of whether or not gestures fall under the co-operative principle or the communicative principle of relevance? In other words, what is it that compels people to pay attention to gesture?

There are at least two potential answers to these questions that have an impact on pragmatic theory. The first, which can be attributed to Enfield (2009b), is that comprehenders pay attention to gesture because gestures are taken as part of composite signals, which are composed of different signals combined to be communicative (often communicating more than the sum of their parts). Enfield (2009b, p. 16–17) states that gestures fall under two heuristics that are involved in the processing of utterances. First, the *contextual association heuristic* states that signs that are associated in terms of proximity and temporality, for example, should be taken as being part of a “single signifying action” (Enfield, 2009b, p. 16). The second heuristic is the *unified utterance-meaning heuristic*, which states that “contextually associated signs point to a unified, single, addressed utterance-meaning” (Enfield, 2009b, p. 17). The alternative view, which can be attributed to Wharton (2003) and Wharton (2009) is that iconic gestures are interpreted because they are taken as natural signs which have been deliberately shown. Wharton builds on the Gricean distinction between meaning and showing. Natural signs, such as coughs, are meaningful because they exist in a contiguous relation with their referents. Since coughing is typically (not conventionally) associated with illness, coughing can be taken as a sign that someone is ill. In the right condition, such signs can be deliberately

shown. According to this view, gestures communicate because they help speakers speak. Such gestures, can then be deliberately shown by the utterance producer. In other words, gestures are not intentionally communicative, but communicate because they communicate naturally. Therefore, Wharton's view acknowledges the first of Enfield's heuristics but not the second. For Enfield, all utterance elements are equally meaningful, however, Wharton's view seems to place primacy to coded linguistic comprehension, suggesting that non-verbal communication is vaguer and weaker (Wharton, 2009, p. 192).

What effect might this distinction have on how comprehenders process gesture? If the unified utterance-meaning heuristic is correct then it would be expected that gesture will necessarily be processed and affect speech comprehension. However, if we follow Wharton's view then the processing of gesture might be expected to be triggered by encoded linguistic meaning. Further, gestural information will only play a role in utterance comprehension when it conveys information not represented in speech. This is because Wharton's view builds on the relevance theoretic comprehension procedure, which can be described as follows: "[t]he hearer takes the conceptual structure constructed by linguistic decoding; following a path of least effort, he enriches this at the explicit level and complements it at the implicit level, until the resulting interpretation meets his expectations of relevance; at which point, he stops" (Sperber and Wilson, 2012, p. 39). The reason why this procedure suggests that gesture which conveys the same content as speech will not be processed is because it is difficult to explain why the gesture is worth processing if it is not part of the utterance producer's informative intention. Moreover, if gestures are only comprehended inferentially, it might be expected that the extraction of meaning from gesture will be slower than the extraction of meaning from linguistic signals.

### 5.1.1 This study

Questions involving the comprehension of speech and gesture cannot be addressed by focussing on level four of interaction action ladder, because it is not possible to analyse the online comprehension of utterances and, as such, only the end product of comprehension is available to the analyst. However, it is possible that the answers to the ques-

tions of real-time speech and gesture integration happen at a sub-action level. In order to capture gesture's effect on real-time comprehension, the experiments reported in this chapter involve participants watching short video clips and selecting an item from a two by two array based on what they have seen and heard the on-screen character refer to. These experiments are designed to mimic the comprehension of utterances used during the map task. However, the end goal for participants in the map task is for the follower to draw a route on their map. For this reason, many utterances in the map task are instructions of what a follower should draw, taking the form of imperatives including the verb "draw". For this reason the videos used in chapter include imperative utterances including the verb draw, however the experimental participants are only required to select the landmark that best fits the utterance in the video. This mismatch in perlocutionary effect between video instructions and experimental instructions was intended to reduce the participant to the status of 'overhearer' and stop them from being required to reach the level 4, proposing and considering, of the action ladder. In other words, participants were required to behave as if they were only interested in the meaning of the words and not the illocution of the speech act. The reason for this is because this study is specifically interested in addressees' comprehension of the semantic content of speech and gesture. The paradigm adopted is the Visual World Paradigm and therefore the next section will present a history of relevant studies and an outline of the methodology before describing some of the considerations that must be taken when using it to investigate co-speech gesture.

## 5.2 The Visual World Paradigm

The visual world paradigm was first devised in 1974 by Roger Cooper. Cooper discovered that people presented with spoken language and a visual field containing elements related to the informative content of speech tend to shift their gaze and fixate on those elements most closely related to the concurrent speech. However, it wasn't until Tanenhaus et al. (1995) that the methodology was recognised as providing crucial insight into online language comprehension with an ability to answer questions regarding the integration of visual and linguistic processing that had evaded psychologists and psycholin-

guists using the lexical decision tasks that were pervasive at the time.

Studies employing the visual world paradigm have largely differed according to two aspects: first, the ‘visual world’ of the visual world paradigm has differed in terms of its composition (see figure 5.1); and, second, in terms of what is expected of the participants. In terms of the visual display, some studies use the space on a tabletop in front of the participant containing real objects with which the participant interacts (Tanenhaus et al., 1995). In other studies employing a physical array is erected in the centre of the table in between the participants (Metzing and Brennan, 2003). The latter allows agents to interact with each other. Very commonly, rather than using a physical space, a digital space displayed on a computer monitor acts as the visual world, which can either be semi-realistic representations (Altmann and Kamide, 1999) or arrays of objects that could include two or more items (Alloppenna, Magnuson, and Tanenhaus, 1998). There are also studies with visual worlds composed of written words instead of items (Salverda and Tanenhaus, 2010) and studies that present a blank screen during the presentation of audio instructions (Altmann, 2004). In terms of what participants are required to do, they may have to move things around a physical array or a digital one (by dragging and dropping), click on objects in a digital array, or simply look at the visual world while they watched and listened a stimulus.

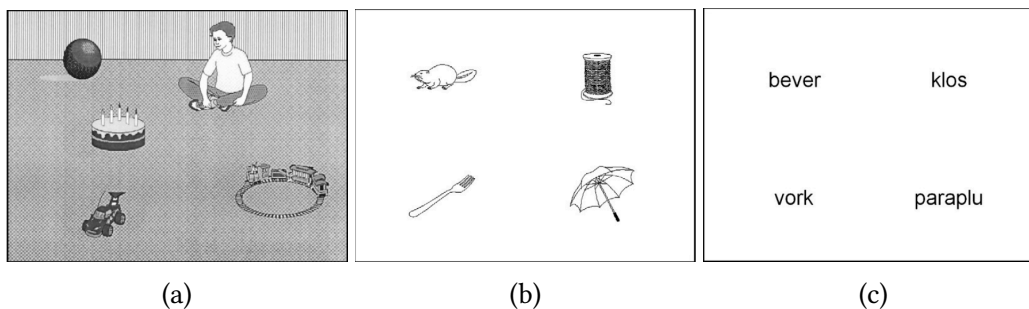


Figure 5.1: Three different digital visual worlds. (5.1a) is taken from Altmann and Kamide (1999) and (5.1b) and (5.1c) are taken from Huettig and McQueen (2007). For a more detailed discussion of these visual worlds see Huettig, Rommers, and Meyer (2011, p. 153).

### 5.2.1 A relevant history of the visual world paradigm

Although the set up of the visual world paradigm remains largely fixed, there has been a great deal of variation in what such studies have investigated. Moreover, the visual world

paradigms has been an extremely influential paradigm in various disciplines. Since the focus of this section is using the visual world paradigm to study speech-and-gesture utterances, this section will focus on four distinct areas:

- Early visual world paradigm
- Sentence processing
- Pragmatics
- Visual Processing

In a series of ground breaking experiments, a team consisting of Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy (cf. Tanenhaus et al., 1995; Eberhard et al., 1995), demonstrated how monitoring eye-movements could be used to answer questions which had eluded linguists for at least half a decade. In the simple version of their experiment, they demonstrated that an individual viewing a scene and hearing an utterance containing one of three modifying adjectival phrases (e.g., starred, red, square) would look at the target earlier when it was disambiguated earlier in the noun phrase than when it was disambiguated later. For example, a participant looking at a scene with one starred item would look at the target object 275ms following the offset of the word “starred”. This is true for the word “red” in scenes containing only one red object and the word “square” in scenes containing one square. Accounting for the fact that it takes roughly 200ms to programme an eye movement (Matin, Shao, and Boff, 1993; Hallett, 1986), this suggests that people comprehend the meaning of a lexical item 75ms after it is heard, thereby providing evidence that the processing of spoken language is incremental and that visual context is rapidly integrated with it (Eberhard et al., 1995, p.417). The team also repeated this finding with more complex displays and utterances to demonstrate that the participants were not simply fixating on the semantic content associated with each lexical item.

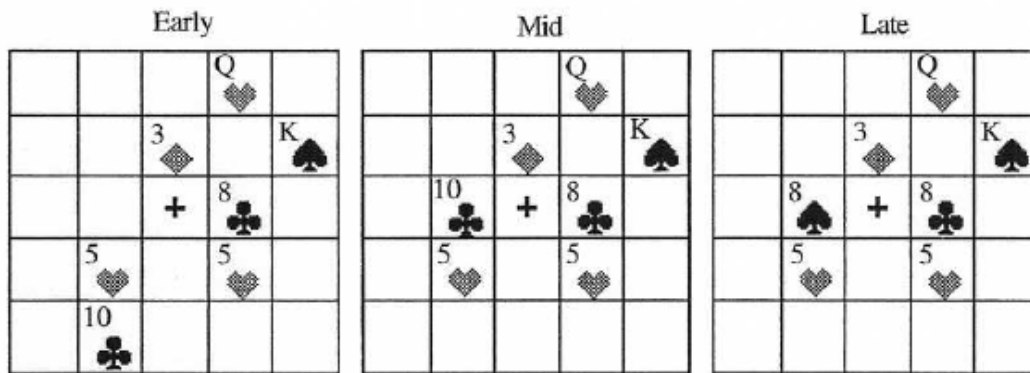


Figure 5.2: Visual world adopted in Eberhard et al. (1995, Experiment 2).

In this experiment (Eberhard et al., 1995, experiment 2, see figure 5.2), scenes containing standard playing cards were shown to participants who would be given instructions such as “Put the five of hearts that is below the eight of clubs above the three of diamonds”. The scenes would contain two five of hearts cards so it would fall to the post-modifying clause to provide a disambiguating function. In three conditions there were different “context” cards, thus in the “early” disambiguation condition there was only a card above one of the five of hearts; in the “mid” condition there was card above both five of hearts but only one was an eight; in the “late” condition there was an eight card above both five of hearts cards but one was the eight of spades and the other was the eight of clubs. Therefore, in this experiment the preposition “above”, number “eight”, and suit “clubs” were being used to disambiguate the target at different points. Once again they found that fixation to target item occurred as an incremental process, with average temporal latency going from early to late (Eberhard et al., 1995).

Another finding (Eberhard et al., 1995), and probably the most significant for theories of syntactic processing, involves an experiment in which participants were asked to move objects using utterances that referred to one object on a table (e.g., “Put the apple that’s on the towel in the box”). The sentence: (1) “Put the apple on the towel in the box” is ambiguous before the phrase (“in the box”) because it is not clear whether the towel represents the apple’s location (“on the towel” as relative clause / location) or its destination (“on the towel” as prepositional phrase / destination). The ambiguity can be highlighted by comparing (1) to (2) “Put the apple that’s on the towel in the box” in which relative clause reading is forced due to the explicit presence of “that’s” and therefore the



prepositional phrase “on the towel” is understood to be the location of the apple rather than the destination. Unlike the other experiments detailed in Eberhard et al. (1995) this ambiguity is syntactically rather than lexically motivated. In the experiment, they compared visual worlds (see 5.3) in which there was either (i) one apple on a towel, (ii) a towel, (iii) a box, and (iv) a pencil or (i) one apple on a towel, (ii) one apple on a handkerchief, (iii) a towel, and (iv) a box. From these visual worlds, it should be clear that in 5.3a the first part of the utterance “Put the apple on the towel” would be more likely to be understood as if “on the towel” is describing the destination, since “the apple” already uniquely specifies the object to be moved. In 5.3b, “apple on the towel” would be used to distinguish between the two apples (i.e., “on the towel” and “on the handkerchief”). Tanenhaus et al. (1995) found that in the condition with one apple, participants looked at the empty towel with “on the towel”, suggesting they are treating it as a destination. However, in the condition with two apples they did not, suggesting that they interpreted it as a modifier describing the location of the object. These findings demonstrate that the processing that syntactic structures are sensitive to contextual information in real time.

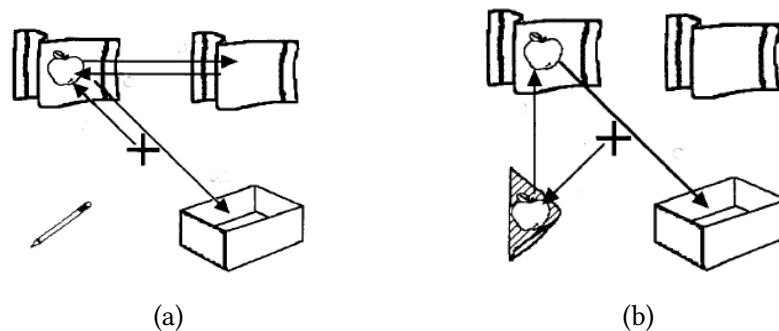


Figure 5.3: Visual worlds from Tanenhaus et al. (1995). (a) depicts the scene with only one apple, for which “on the towel” would describe the destination. (b) depicts the scene with two apples, for which “on the towel” describes the apples current location.

Allopenna, Magnuson, and Tanenhaus (1998) replicated many of the findings listed above but used a digital monitor to represent the visual world. Following the work of Tanenhaus et al. (1995) has been two decades of research employing the visual world paradigm. Those studies that are relevant to the current chapter can be broken into semantic processing, informativity, and visual processing (for a review, see Huettig, Rommers, and Meyer, 2011).

### 5.2.2 Semantic Processing

Regarding semantic processing, several studies have demonstrated findings beyond those of Tanenhaus et al. (1995). In a striking study, Chambers, Tanenhaus, and Magnuson (2004) found that participants faced with an ambiguous sentence such as “Pour the eggs in the bowl over the flour” and a scene depicting either (i) a bowl, (ii) some flour, (iii) a broken egg in the bowl, and (iv) a broken egg in a glass or an identical scene except the egg in the glass is whole, would anticipatorily fixate on the bowl as a goal in the condition when the egg in the glass is whole. This is because the whole egg is not a suitable object for the verb “pour”. Moreover, in the condition with the broken egg in both vessels this effect is not observed. This study further suggests an ongoing interaction between sentence and visual processing over the time course of comprehension. Further, in a series of studies, Altmann, Kamide, and colleagues (Altmann and Mirković, 2009; Altmann and Kamide, 2009; Altmann and Kamide, 2007; Altmann, 2004; Altmann and Kamide, 1999) have demonstrated the importance of predictive fixations as a measure of expectancy. Altmann and Kamide (1999) use a semi-realistic cartoon scene containing an image of a boy, a ball, a cake, a toy car, and a train set (see 5.1a) to investigate the time course of eye movements to target following an utterance of “the boy will eat the cake” or “the boy will move the cake”. They found that participants would make saccadic movements to the target (“the cake”) before the determiner of the verbal complement in the “eat” condition but not in the “move” condition, demonstrating that eye movements are predictive of the upcoming information when the syntactic context makes one item in an array more likely to be the verbal complement. Altmann (2004) reported a similar effect, however, this time the content of the visual world was replaced with a blank screen. His findings suggest that participants made the same predictive eye movement but to the location where the target would have been on the screen, demonstrating that eye movements are not dependent on the concurrent visual scene, but are dependent on a “mental record” of the scene. In a further development, Altmann and Kamide (2007) demonstrated that the tense of the verb can result in different anticipatory fixations. When participants heard either “The man will drink [...]” or “The man has drunk [...]” when faced with a visual world containing both an empty wine glass and a full glass of

beer, it was demonstrated that the past tense “drunk” results in looks to the empty glass whereas present tense “drink” produces looks to the beer. These findings are important to this thesis because they demonstrate that features of particular lexical items can have an effect on the fixations associated with utterance processing.

### 5.2.3 Informativity

Engelhardt, Bailey, and Ferreira (2006, experiment 3) investigated the effects of over- and under-informative utterances, based on the Gricean maxim of quantity (Grice, 1975), which states that speakers should be as informative as required and no more informative than is required. In their third experiment, they explored how participants reacted to over and under informativeness by analysing fixations to the targets following instructions to move items around an array. Instructions came in one of four conditions, based on two variables. These were whether the location matched the current location of the object and whether the utterance includes a modifying prepositional phrase. So, if the apple was already on a towel the four conditions would be: (i) + match, - post-modifier (“Put the apple on the towel”); (ii) - match, - post-modifier (“Put the apple in the box”); (iii) + match, + post-modifier (“Put the apple on the towel on the other towel); and (iv) - match, + post-modifier (“Put the apple on the towel in the box”). The display consisted of four objects: (i) an apple on a towel; (ii) a towel; (iii) a frog; and (iv) a box. Therefore, the variables manipulated were: (a) whether or not the location to which the objects were to be moved matched or did not match the current location of the object, and (b) whether or not the object was referred to using a post-modifying prepositional phrase. Engelhardt, Bailey, and Ferreira (2006, p. 568) found that in the condition where participants were given “on the towel” as post modifier of “the apple” followed by the goal “in the box” there was a delay in fixating on the box, which Engelhardt, Bailey, and Ferreira (2006) take to represent a momentary confusion. This finding suggests that participants in the visual world paradigm do not anticipate instructions to provide more information than is necessary, suggesting that hearers expect speakers not to be over-informative. These findings are relevant to the current study because if speech and gesture are both taken to convey the same information (and they are both taken to be intentionally produced

by the utterance producer), then it might be expected that addressees will behave in a similar way to when speakers are over-informative.

#### 5.2.4 Visual World and Visual processing

Gesture is inherently visual, therefore visual processing is of particular interest in this study. Therefore, a related question facilitated by visual world paradigm is whether or not comprehenders look at items that share visual characteristics with a named object. For example, Dahan and Tanenhaus (2005) investigate the probability of looking at a rope (visual competitor) upon hearing the word “*snake*”, and whether or not there was an increase in fixations to the rope over the other two unrelated distractors (a picture of a couch and an umbrella). Importantly, target and visual competitor, while being visually related, were not semantically related. This study aims to clear up several ambiguities relating to previous visual world studies. For example, although it has been demonstrated that participants fixate on target items when they hear a word that can be used to refer to that item, it is not clear if this is because the image in the visual world interacts with the linguistic information associated with the word or whether there is some implicit phonological activation based on viewing the scene. In the latter case it would be a case of matching the implicitly generated linguistic representation (e.g., representing the word “*snake*” upon seeing a picture of a snake) with the one explicitly generated via the utterance (representing the word “*snake*” after hearing the word “*snake*”). As Dahan and Tanenhaus (2005, p. 454) argue “such a visual competition effect could be explained in terms of a match between the conceptual and visual representations associated with the unfolding spoken word and a coarse structural representation of the competitor picture”. To combat this issue, Dahan and Tanenhaus (2005) included two different preview lengths (300ms and 1,000ms). They argue that if heightened fixations to target are based on implicit phonological activation, then increasing the time window before a target word is heard should result in a decrease in looks to the visual competitor. The reasoning behind this is that if there is a visual competition effect, then providing participants with a longer time period to explore the visual world should result in a higher activation of phonological features of the items. Thus, the 300ms condition should produce more

looks to the visual competitor than the 1,000ms condition. As Dahan and Tanenhaus (2005, p. 454) argue “such a visual competition effect could be explained in terms of a match between the conceptual and visual representations associated with the unfolding spoken word and a coarse structural representation of the competitor picture”.

Dahan and Tanenhaus’s 2005 results show that not only is there a clear visual competition effect, with participants being statistically more likely to fixate on the visual competitor than any other distractor, but also this effect was greater in the 1,000 ms condition. This result is the opposite of what one would expect on the assumption that the visual items in the display produce implicit naming, thereby generating phonological representations. Rather, their findings suggest that top-down target representations generated via linguistic input are mapped onto the bottom-up representation of the visual world.

Huettig and Altmann (2007) challenge the account of visual competitors in Dahan and Tanenhaus (2005). Instead, Huettig and Altmann (2007, p. 989) argue for an account which takes language representations as “modulating the activation of (already activated) conceptual visual display-based representations”. Such “display-based” representation would be activated by viewing the scene before the linguistic input is comprehended. They argue that because Dahan and Tanenhaus (2005) use a visual world that contains both the linguistic target and the visual competitor at the same time, they cannot generalise their findings to all visual world paradigms. The main suggestion is that while those visual world experiments involving target selection (e.g., via mouse click) may involve an element of visual search which requires the top down processing of linguistic input and bottom up processing of visual information, in other versions of the paradigm, specifically ones in which the participant acts as a passive observer, the opposite might be the case. In other words, the representations derived from the images in the array are not exclusively related to the token images in the scene but to the types of objects they are related to. For example, an image of a snake is taken as an instance of an image of a more general visual snake representation.

To test these ideas, Dahan and Tanenhaus (2005) employ biasing contexts where a mentioned element early in a sentence contextually limits the possible potential refer-

ents. For example, the word “zookeeper” would increase the contextual suitability of a snake when accompanied by a scene containing a snake, a barrel, a pillow, and a carpet. This would in turn lead to the snake being the most salient item in the display even before it is referred to. If this is the case then it should lead to more looks to the snake prior to the offset of the word “zookeeper”. The study focussed on whether visually similar items (e.g., cable) would be fixated on in a snake-biasing context. This would suggest that the fixations on the cable are due to top-down processes related to the conceptual activation of snake via “zookeeper”. Their argument is that as the participants view the scene (prior to linguistic input) they will generate conceptual visual display-based representations. For example, if a scene contains a snake this will lead to an activation of the representation of a snake, which may then be further boosted due to the semantic links to the word “zookeeper”. However, in the scene containing the visual competitor (e.g., cable (analogous to rope in the Dahan and Tanenhaus (2005) case)) then there should be no early fixations on the visual competitor because there is no semantic link between the visual competitor and the biasing device to boost the activation of the visual competitor. This study therefore attempts to delve deeper into the specifics of why visual competitors are fixated on.

Huettig and Altmann (2007) explored these ideas in two experiments. In experiment 1, they investigated whether or not the lexical item referred to increases fixations on another item that is visually but not semantically related. Participants heard one of two utterances:

1. *In the beginning, the man watched closely, but then he looked at the snake and realized it was harmless*
2. *In the beginning, the zookeeper worried greatly, but he looked at the snake and realized that it was harmless*

The first utterance does not increase the likelihood that the snake will be referred to whereas the second one does. Different scenes were used which either contained the picture of a snake or the picture of a cable (which is visually similar to a snake). From this, Huettig and Altmann (2007) produced three conditions: (i) *Neutral* in which utterance

(1) is produced while participants view the scene containing the snake (ii) *Biasing* in which utterance (2) is produced while participants view the scene containing the snake and (iii) *Competitor* in which utterance (2) is produced while the participants view the scene containing the cable.

Huettig and Altmann (2007) explored the proportion of fixations within three time windows: (1) the onset of the target word (i.e., “snake”), (2) the offset of the target word, (3) and 200 ms after the offset of the target word. They found that participants were significantly more likely to fixate on the target in the biasing condition at the onset of the target word (*snake*) than any other item in the scene but not in the neutral condition. And, more interestingly, they found that participants were more likely to look at either the snake or the cable at the offset and 200 ms after the offset in all conditions. The findings from Huettig and Altmann (2007) experiment 1 show that people may fixate on visually similar objects even when the object explicitly referred to in the utterance is not present in the scene. However, looks to the visual competitor are not biased by the preceding linguistic contexts, whereas looks to the target are. Huettig and Altmann (2007, p. 1002) use this as evidence to argue against “an explanation of this pretarget word bias in terms of “*zookeeper*” causing the activation of shape representations associated with snakes which are then matched against visual form information extracted directly from the image”. Their conclusion is that the higher probability of early looks towards the snake in the biasing condition is due to an episodic representation of the depicted snake being related conceptually to the conceptual representations associated with zookeepers. Furthermore, this episodic representation is additionally activated upon hearing the word “zookeeper”. This roundabout way of explaining visual similarity effects captures the preference to fixate on the snake but not the cable since the cable is only visually connected to snake and does not have a conceptual link to zookeeper.

In their second experiment, Huettig and Altmann (2007) test whether or not the visual competitor effect can be demonstrated in a situation where the critical word used to refer to the object in the scene is a homonym with a dominant and subordinate meaning. For example, “pen” can be used to refer to both a writing implement and a cage in which animals are kept. Using a word association task, Huettig and Altmann (2007, p. 1005)

determine which meaning of the homonym is dominant and which is subordinate. For example, a participant in the norming study would be given the word “Boxer” and they might say “fighter” or they might say “dog”. Similarly, in the case of “pen” they might say “pencil” or they might say “pig”. For experiment 2, all biasing conditions increase the salience of the subordinate meaning. However, this time the visual competitor will only be related to the dominant meaning. In the case of “pen” the image representing the visual competitor was a needle.

The procedure for the experiment was the same as experiment 1, but with different utterances and scenes. The utterances would be in one of two forms:

1. “*First, the man got ready quickly, but then he checked the pen and suspected that it was damaged*”
2. “*First, the welder locked up carefully, but then he checked the pen and suspected it was damaged*”

These can then be used to create the three conditions by presenting them with different visual scenes. In the *neutral* condition, participants would hear sentence 1 and see a scene containing: (i) a pen (writing implement), (ii) a pen (cage), (iii) a bicycle, and (iv) a bucket. In the *biasing* condition, the scene would contain the same items as in the *neutral* condition except this time they heard sentence 2. And in the *competitor* condition, they heard sentence 2 but the scene contained a needle instead of a pen (writing implement). The *neutral* condition is intended not to bias either reading of the word. The *biasing* and *competitor* conditions were designed to bias the subordinate reading of the word.

To summarise the findings of this experiment, Huettig and Altmann (2007) found that at the onset of the target “pen” participants were more likely to fixate on the subordinate referent in the *biasing* and *competitor* conditions. In the *neutral* condition, participants were more likely to look at the dominant referent at both the offset and 200ms following the offset of the target word (they were also significantly more likely to look subordinate referent than the two distractors at both these time points). In the *biasing* condition, at offset and 200ms after offset time points, this relationship was reversed so that participants were most likely to look at the subordinate referent, but they were still significantly more likely to look at the dominant referent than any of the distractors. Lastly,



this pattern is (almost identically) replicated in the *competitor* condition with the shape competitor taking the place of the dominant referent. As Huettig and Altmann (2007) note, the key finding of this experiment was that in the competitor condition participants still directed visual attention towards a visual competitor of a dominant homonym. This was the case even though there was a picture of the subordinate homonym present in the scene and the linguistic context strongly biased towards the subordinate meaning of the homonym. This demonstrates that even though it was contextually non-salient the dominant meaning of the homonym is still activated along with perceptual information associated with it.

Huettig and Altmann (2007, p. 1014) suggest the following outline from task performance: (1) the display starts and the participant views the four objects (2) picture-derived activations, including visual form representation and spatial location information are created (3) spoken linguistic input creates language-derived representation (once again including visual form representations) are activated (4) overlap between the visual form representations activated via both the picture-derived representations and language-derived representations increases the number of eye movements directed towards the competitor objects. The critical point to their explanation is that visual form representations are **reactivated** rather than simply activated by the linguistic input. These observations are of crucial importance for the study of gesture because they demonstrate that the processing of visual information plays a crucial role in the increased fixations to the target. Gesture, which, has been said to “promote image-based simulations of the meaning of an utterance” (Wu and Coulson, 2014, p. 49) may facilitate the activation of the target within the array. If this is the case then the visual world paradigm could offer a crucial insight into the real-time processing of speech and gesture.

### 5.2.5 Visual world and gesture

To my knowledge, Silverman et al. (2010) is the only published study which uses the visual world paradigm to study the online comprehension process associated with co-speech gesture. The focus of the study is on the processing of gesture by individuals with autism, and age, gender, Verbal IQ, and socio-economic status matched controls.

Since there is so little written on the effect of gesture as measured by the visual world paradigm all the findings reported within Silverman et al. (2010).

The study rests on the assumption that if iconic (gestural) information is integrated with semantic (spoken) information from a previous part of an utterance, then comprehenders will fixate on a target object sooner when gesture is presented than when speech occurs alone (Silverman et al., 2010, p. 382). If this assumption holds then their study allows Silverman et al. (2010) to measure whether and to what extent gesture facilitates reference resolution. Silverman et al. (2010) predict that for the typical controls gesture will facilitate reference resolution resulting in reduced latencies for target fixations. Further, they predict that this will not be the case for the group with autism, due to evidence that shows individuals with autism have difficulty integrating information across auditory and visual modalities (Silverman et al., 2010, p. 382). Additionally, they do not predict that these integration difficulties are the result of processing gestural information alone but gestural information that is concurrent with speech. In other words, integration and not necessarily modality, is the focus of the study.

To test these predictions, Silverman et al. (2010) conducted two experimental tasks. The first task included two conditions: speech only and speech-and-gesture. In the speech only condition referring expressions were produced without concurrent gesture, whereas in the speech-and-gesture condition, speech was accompanied by a gesture. The display contained complex shapes consisting of two components: 1. a *nameable* component, which is easy to name but difficult to gesture (e.g., a mitten) and 2. a *gesturable* component, which is easy to depict through gesture but not to name (e.g., a line with two loops). The second task constituted a gesture only condition. In the gesture only condition, referring expressions were identical to those in the speech-and-gesture without the speech. For this task, the arrays only included the gestural component. This control condition was used to test whether gestural information could be comprehended. Participants viewed arrays and were asked to select the item being depicted. Both groups performed close to ceiling on the control task, demonstrating that they can process gestures (Silverman et al., 2010, p. 384).

In the speech only and speech-and-gesture conditions the set up was a bit different.

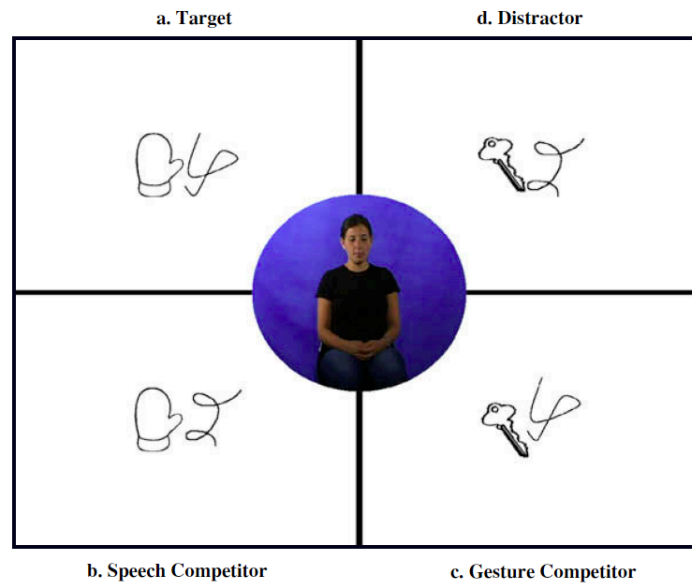


Figure 5.4: Example of Array from Silverman et al. (2010, p. 384)

For these conditions the arrays contained four complex shapes, one of which was the target. The complex shapes were arranged so that they were related to one another in various ways. The target contained both a gestural component and a nameable component that were both related to the information conveyed in a video depicting gesturer’s upper body (see 5.4). The non-target objects formed three separate groups:

1. *Speech competitor*, which shared the nameable component of the target
2. *Gesture competitor*, which shared the gesturable component of the target
3. *Distractor*, which didn’t share either component with the target

Figure 5.4 shows an example display. The accompanying utterance is “A mitten and a line with one loop through it.” and in the speech only trial the model would have left her hands on her lap, while in the speech-and-gesture trials she gestured *one loop* by tracing the outline of the loop with her index finger. Importantly, the stroke phase of the *one loop* gesture slightly preceded the word “one”. In this way, target disambiguation occurs earlier in the speech-and-gesture trials than in the speech only ones.

Silverman et al. (2010) calculated the average proportion of fixations to each spatial area of interest within three temporal regions, with 0ms point representing the *point*

of *disambiguation* (POD). Region 1 is 800ms before to 200ms after the speech POD; region 2 is 200ms-1200ms after speech POD; and Region 3 is 1200-2200ms after speech POD. The fixations were coded as belonging to six areas of interest: (1) target, (2) speech competitor, (3) gesture competitor, (4) distractor, (5) video of the model, (6) and excluded (i.e., blinks, saccadic movements, fixations between items, and off-screen fixations). In the critical analysis, Silverman et al. (2010) analysed whether participants fixated on the target compared to the speech competitor more quickly in the speech-and-gesture condition vs the speech only condition. They did this by calculating a target ratio score, which was derived by computing the ratio between the proportion of fixations on the target and the sum of the proportion of fixations on the target and the sum of the proportion of fixations on the target and speech competitor. This target ratio score was compared to chance (.50) to investigate whether or not there was a preference to fixate on the target. For region one, they found that both groups were no more likely to look at the target than the speech competitor in both conditions. In contrast, for region three, they found that both groups were more likely to fixate on the target than the speech competitor for both conditions. In the critical region, region 2, they found different results for the two groups. In the speech only condition, the control group were no more likely to look at the target than the speech competitor whereas the autism group were more likely to fixate on the target. However, in the speech-and-gesture condition the control group fixated significantly more on the target whereas the autism group did not. In a separate analysis, Silverman et al. (2010) also calculated reaction times based on mouse clicks on the target object. Lastly, they also demonstrate that participants from both groups are more likely to look at the video in the speech-and-gesture condition than in the speech only condition.

To summarise their findings, Silverman et al. (2010) found that both groups are able to understand the meaning of gestures in the gesture only condition. In the speech only and speech-and-gesture condition, both groups are eventually able to fixate on the target (based on region 3). However, in the speech-and-gesture condition, controls are more likely to fixate on the target during region 2 whereas the group with autism were not. This suggests that for controls, gesture information is immediately integrated into the

processing of speech whereas for the autism group this is not the case. In relation to their predictions, these results demonstrate that controls can and do integrate gesture in real-time, whereas individuals with autism do not. However, it is important to highlight that they do not find a significant difference in terms of reaction time. In other words, the difference seems to be an effect of real-time processing. This further emphasises the need for studies of the real-time processing of gesture over those that explore gesture's effect within an interactive setting.

The findings of Silverman et al. (2010) are comparable to those found in the neuroscientific explorations of gesture, which were outlined above (Section 2.2.4). Özyürek (2014, p.8) summarises these findings as suggesting that in terms of comprehension, gestures “are not perceived as mere incidental accompaniments to speech”. Such gestures are processed “semantically” during comprehension as “integrated part of the speaker's communicative message”. What's more, gestures are not automatically processed but, comprehenders take the communicative intent of the speaker into account (Özyürek, 2014).

Taking the above studies together, the visual world paradigm is a tool that can be used to explore the real-time comprehension of utterances in relation to visual context. It has been shown that comprehenders are sensitive to ambiguities that arise because of syntactic and semantic contributions to meaning. What's more, over-informativeness can result in increased fixation latencies to a target object. In terms of visual processing, it was demonstrated that comprehenders are sensitive not just to the linguistic meaning of the utterance but to the spatial representations evoked by linguistic realisations. Finally, it has been shown that gesture can be integrated in real-time. These findings all point to the fact that the visual world paradigm has potential as a tool for accessing the real-time comprehension of speech-and-gesture utterances. What's more, the visual world paradigm provides the opportunity for the analyst to explore time regions in between the point in which the target item is disambiguated but before a selection (via a mouse click) has been made. This time window in between disambiguation and selection is crucial for understanding the process leading up to what Clark (1996) calls *recognising*, which is the point at which an individual has recognised the signal that an

interlocutor produced.

### 5.3 The current study

This study builds on several questions raised by Silverman et al. (2010). Their study investigated whether or not gestured information has an effect on comprehension, but it only does so in a condition for which the information is also presented through speech, since their gesture only condition is a control condition technically from a different experiment that used a different array. Additionally, gesture is presented as an all or nothing feature of communication which either delivers analogous information to speech, or nothing. This is far from the everyday composition of speech and gesture, where gesture often complements the semantic content of speech (Kendon, 2004, pp. 176–177). Further, the complex shapes used by Silverman et al. (2010) consist of two elements that are put together with no real reason for them to naturally form a pair. Therefore, the gesture is actually referring to an entire element of the complex shape and the speech is being used to refer to another. Third, the time between gesture POD and speech POD represents a short time window, making it difficult to fully explore the effect of gesture as a disambiguating aspect of referring expressions.

This current study explores the effect of gesture on the comprehension of utterances containing spatial descriptions, which were designed to be similar to those found in the map task (see chapters 3 and 4 above). In order to achieve this, the study employs an elaboration of the experiment in Silverman et al. (2010) with some important modifications. In particular, the study focuses on a specific local semantic element of the utterance (in the sense of Beattie and Shovelton (2006)) which is used to refer to a particular element of a composite image, in this case the manner of the route, as opposed to a whole “gesturable” element of a complex item. For example, De Ruiter (2007) describes a gesture produced in the narration of a scene from a cartoon. The gesture depicted the movement of a cartoon cat, who was being propelled by a bowling ball that had just been dropped down his throat, as he travels down an alley. Importantly, the gesture does not represent the whole scene (including background objects), nor does it even represent the whole cat. The gesture only depicts the *manner* and *path* of the cat. It was these features that

the narrator (in de Ruiter's 2007 example) focussed on with gesture and not the whole scene.

### 5.3.1 Semantics of spatial descriptions

The semantic features used to create the stimuli in this study are taken from the categories used to analyse the map task in chapters 3 and 4. Taking a typical utterance (e.g., “and you're gonna loop round the right of the pyramid”), one finds in the map task it is possible to generalise and suggest that such utterances frequently include four semantic elements:

1. **GROUND:** a landmark used to anchor the description of the route. In speech this would be provided by the noun associated with the landmark (e.g., “broken gate”) whereas in gesture it is only really described through deictic gestures that represent the landmark's position in space.
2. **ORIENTATION:** the relationship between the route and the ground. In spoken English this is expressed with terms such as “above” and “to the right of” and in gesture it is typically expressed through the relationship of the dominant and non-dominant hands, where the non-dominant hand might represent the ground (as described above) while the dominant hand traces the route, typically with the index finger.
3. **DIRECTION:** the direction in which the route is said to travel. In spoken English, it is expressed through verb + prepositional adjuncts such as “going leftward”, “travelling downward”. In gesture, it is depicted through the direction of movement of the hand in relation to the speaker's body and the observers perspective.
4. **MANNER:** the shape of the path. In spoken English it is expressed through adjectives such as “wiggly” or “zigzag”. In gesture, manner is usually realised in conjunction with direction and is depicted through small alterations in movement. For example a rightward movement may be made up of jagged alterations (depicting a zig-zagging line) or smooth, random alterations (depicting a wiggly line).

For the task reported here, these semantic elements are used to create both the composite image that fill the array and the utterances that form the instructions. The next two sections will cover the creation of each in turn.

### 5.3.2 Creating array items

The visual world used in this study consisted of a 2x2 array with a video of a head-and-torso model producing speech-gesture composites in the centre (see figure 5.6). The four items in the array were images composed of two parts. The central images of the items were taken from the subset of images from the HCRC map task corpus used in the map task described in the previous chapters (cf. Anderson et al., 1991). These images consisted of a monochrome picture and a label written in English. The image forms the component of the items which is easier to describe through speech than gesture. Five different images were used: (1) Camera Shop (2) Broken Gate (3) Lemon Grove (4) Slate Mountain and (5) Pyramid.

The second component, which is a dashed line, can best be described by imagining a horizontal and vertical axis crossing through the centre of the image. This dashed line, which will be referred to as the route, forms a path which either runs perpendicular to the vertical or horizontal axis or runs around the image crossing the vertical axis twice (above and below the image) and the horizontal axis once. The former of these routes may fall on any of the sides of the image and the latter may cross the horizontal axis on either side. This produces the six different routes represented in Figure 5.5 and thus forms six different ORIENTATIONS: 1. on the left, 2. on the right, 3. above, 4. under, 5. right around, and 6. left around. In addition to the orientation of the routes to the image, they can also differ in MANNER as it will be referred to here. The different *manners* are based on the different spatial adjectives available in English to describe the manner of paths. These are: 1. wiggly, 2. zig-zag, 3. looping, 4. and curving and each can be seen in figure 5.5. A final feature, which is not represented in the image of the path but is nonetheless a necessary component of it, is DIRECTION. DIRECTION appears in pairs of identical images and can be labelled as: 1. going up and going down, 2. going left and going right. In other words, as discussed in chapter 4 above, the route is treated as if



it is in motion, travelling in a certain direction. Literally, however, the route depicted in the image is static, and therefore the image of the route travelling upwards and the image of the route travelling downwards are identical. Together, these features make it possible to produce 480 different images. For the experiment a subset of 40 were used, which were created by producing a collection of composite images with two tokens of each MANNER for each landmark but not having any MANNER/ORIENTATION configuration appear identically with two different landmarks. For example, there would be two lemon groves with wiggly lines and two camera shops with wiggly lines but the orientations for these will necessarily be different (See appendix for a full list of composite images and the encoder script). The lines themselves were created using the Sketchbook application for Android by Autodesk Inc., which ensured that each route element was identical across conditions.

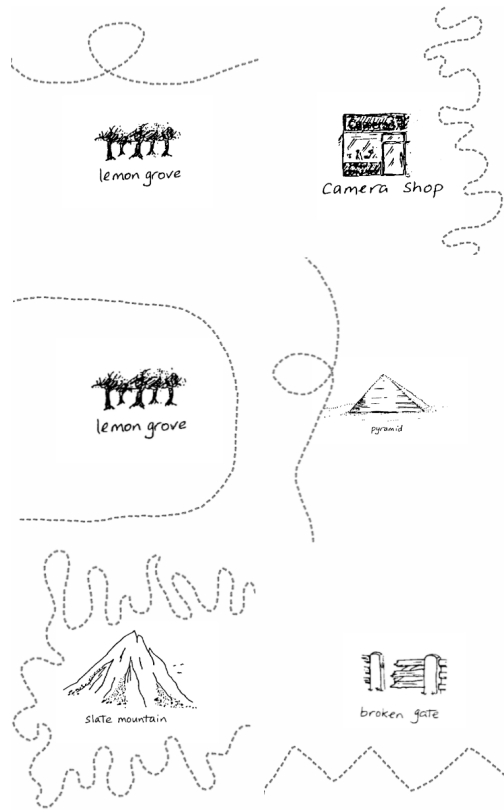


Figure 5.5: Items used to create arrays

The image and the route elements, taken together, are designed to look like a possible section of the maps used in the map task. The *complex* shapes used in Silverman et al. (2010) include two components for which there is no real-world motivation for their pairing (e.g., there is no reason why a mitten would occur with a looped line). However, the items used in the current experiment depict *composite* images because they form naturally occurring pairs (e.g., a route shape and a landmark). The arrays (see figure 5.6) were created by including three other composite images. As can be seen in figure 5.6, two items are related to the target image in terms of direction and orientation and the third is unrelated in terms of all four elements. For example, if the target item was the composite image depicted by a wiggly line under the camera shop (top right), then the routes in the image with the broken gate and the image with the pyramid only differ in

terms of MANNER (the routes are zig-zag and curving, respectively). Every image has a unique landmark.

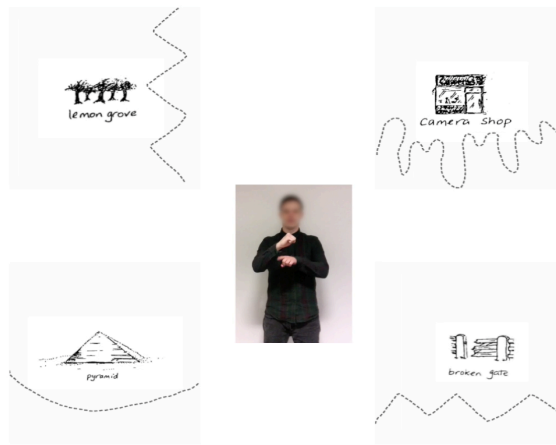


Figure 5.6: Array example

### 5.3.3 Utterances

The utterances, which were designed to accompany the composite images, were created by using the following schema based on semantic analysis of space described in section 3.3.3. The schema captures the distinct elements described in section 5.3.1 and is as follows:

Component	
LANDMARK/GROUND	The nameable component is the image described above which has been taken from the map task. Each nameable component includes its name written underneath it.
ORIENTATION	relates to the relationship between the two elements in the display, which in this case are the landmark and the route shape and can be referred to using expressions such as: “above”, “under”, “right of”, “left of”, and “around”.

DIRECTION	relates to the route shape trajectory and is encoded through a preposition plus the word “going”, i.e., “going right”, “going left”, “going up”, and “going down”. (It is important to note that DIRECTION is not an intrinsic property that can be ascribed to the static images of this experiment, however, because the route shapes are typically described as being representations of movement then DIRECTION is a necessary component of the speech).
MANNER	relates to the shape of the route and is referred to using the spatial adjectives: “wiggly”, “zig-zag”, “curving”, and “looping”). MANNER is the critical element and is manipulated to create the four different conditions. The no manner speech elements are created by placing the word “dotted” in the MANNER slot.

All utterances began with “Draw a” in order to distance the MANNER element from the very beginning of the utterance and to turn the descriptive content of the utterance into an imperative similar to those found in the map task. This was done so that the MANNER component is not the first thing that participants hear. This also means that the participant is not the recipient of the project proposed by the utterer, but is an overhearer meaning that they are only asked to attend to the meaning of the signal and not to goal of the utterer (i.e., getting the comprehender to draw something (Schober and Clark, 1989; Bangerter and Clark, 2003)).

The gestural element of the composite utterance was created by instructing the encoder to use one hand with a closed fist (ASL S handshape) to represent the landmark and the other hand to trace the route with the index finger (ASL G handshape) (see chapter 3 for representations of handshapes). The encoder was instructed that they must convey ORIENTATION, DIRECTION, and MANNER information and shown how each of these elements related to their lexical affiliates in the speech (see encoder materials in

the appendix). However, the encoder was not given more specific information regarding how exactly to trace the route, but offered several opportunities to practice. Four conditions were recorded, two with MANNER information and two without. Furthermore, the encoder was not informed what the specific interests of the study were. For the conditions in which MANNER was not present in the gesture, the encoder was instructed to produce a similar gesture to the MANNER condition except that they should use a flat hand (similar to an ASL B but with the thumb extended). The motivation behind the flat hand is due to the highly conventionalised use of the G handshape for pointing and tracing in the English speaking world (Streeck, 2008).

Crossing the speech and gesture together creates four conditions, all of which relate to MANNER.

Utterance elements	Presence/Absence of MANNER (+/-)
<b>condition 1: speech &amp; gesture</b>	
<i>Speech</i>	+
<i>Gesture</i>	+
Example	
<i>Speech</i>	Draw a wiggly line going right under the camera shop
<i>Gesture</i>	WIGGLY RIGHT UNDER
<b>condition 2: Speech</b>	
<i>Speech</i>	+
<i>Gesture</i>	-
Example	
<i>Speech</i>	Draw a wiggly line going right under the lemon grove
<i>Gesture</i>	NULL RIGHT UNDER
<b>condition 3: Gesture</b>	
<i>Speech</i>	-
<i>Gesture</i>	+
Example	
<i>Speech</i>	Draw a dotted line going right under the lemon grove
<i>Gesture</i>	WIGGLY RIGHT UNDER

<b>condition 4: Neither</b>	
<i>Speech</i>	-
<i>Gesture</i>	-
Example	
<i>Speech</i>	Draw a dotted line going right under the lemon grove
<i>Gesture</i>	NULL RIGHT UNDER

### 5.3.4 Recording and Stimuli creation

The recordings were made in two separate occasions and in blocks determined by the landmark components. Using a Canon Legria mini HD camcorder and audio was recorded separately with a Zoom H2n Audio Recorder. Each condition was recorded separately, meaning that there were two video and audio recordings for each manner condition (+ and -) but only one audio track and one video track for the manner elements were used for the experimental stimuli. For example, because condition 1 and condition 3 include the same + manner in gesture element, but were recorded separately, it could be the case that slight differences in these recordings may have a confounding effect on the experiment. To avoid this, the only one +MANNER in gesture video was used but different audio elements (+/- MANNER in speech) were overlaid. In addition, the audio and video components were shifted so that the onset of the manner component in speech coincided with the onset of the stroke phase of the gesture phrase. To do this all four conditions were compared for how closely speech and gesture components were aligned temporally, and then the most closely aligned were chosen. Therefore, the final videos often had audio out of sync or incongruent with the lip movements. For example, it is possible that the lip pattern was consistent with the word “wiggly” and the audio element included the word “dotted” or vice versa). To overcome this, the encoder’s face was blurred out using an oval overlay, which did not obscure the gesture<sup>1</sup>. For use in the experiment, all recordings were processed using Mac OS X Final Cut Pro.

The arrays were added to the video file in Final Cut Pro, by importing them as .jpegs.

<sup>1</sup>If a participant enquired about the blurring, then they were told that it was motivated by privacy concerns.

Their size was controlled on a pixel by pixel basis to ensure that the composite images were in identical positions across all items. The same was done to the video of the encoder, which was cropped to fit in the space in the centre of the array and controlled so that all video elements were the same size across all items. Initially, videos were placed so as to match the conventional practice of producing spatial information in gesture from the producer's perspective, such that left relates to the producer's left. However, after trialling some of the items, it was decided that the close proximity between the video and the target seemed to increase the effect of the discrepancy between producer and comprehender spatial perspectives. Therefore, all video elements were inverted to be in the comprehender's (i.e., participant's) perspective.

Each video lasted ten seconds, with the video element starting after 4 seconds and the onset of the word "draw" occurred at 5 seconds. In doing so, participants were given enough time to properly acquaint themselves with the items in the array because they would be required to fixate on the video element once it begins (cf. Huettig and Altmann, 2007), although participants were not explicitly told that they had to look at the video element.

### 5.3.5 Points of disambiguation

During the video there are two points of disambiguation (PODs) and these were different depending on the condition. The target object is disambiguated by either MANNER or GROUND, these will be referred to as the MANNER POD and the GROUND POD respectively. The MANNER component appears at the beginning of the utterance and the GROUND component appears at the end. The DIRECTION and ORIENTATION elements will also serve a disambiguating function, but they will only serve to isolate the unrelated item and not the two items that match the target in terms of DIRECTION and ORIENTATION. However, due to the composition of the gestural components, all of the semantic elements in gesture are produced simultaneously and thus DIRECTION and ORIENTATION occur earlier in gesture than they do in speech.

Therefore, taken in conjunction with the array in figure 5.5, the following utterances will have different points of disambiguation marked in **bold**.

*Condition 1: Speech & Gesture*

Draw a **wiggly** line going right under **the camera shop**

WIGGLY

RIGHT

UNDER

*Condition 2: Speech*

Draw a **wiggly** line going right under **the camera shop**

NULL

RIGHT

UNDER

*Condition 3: Gesture*

Draw a dotted line going right under **the camera shop**

WIGGLY

RIGHT

UNDER

*Condition 4: Neither*

Draw a dotted line going right under **the camera shop**

NULL

RIGHT

UNDER

Referring back to figure 5.6, in condition 1 speech and gesture will simultaneously disambiguate the target object (top right in the array). In condition 2, only speech disambiguates the target object, whereas, in condition 3, only gesture does. In condition 4, the object is not disambiguated until the GROUND element is described in speech, which, in this case, is “the camera shop”.



### 5.3.6 Conditions and Variables

The different arrays and utterances generated 160 items (40 arrays by 4 utterance types).

These features generate the following variables in a 2x2 within subjects design:

Independent Variables

Condition		
<i>Speech condition</i>	+	-
Example	“wiggly”	“dotted”
<i>Gesture condition</i>	+	-
Example	WIGGLY	NULL

Dependent Variables

Proportion of eye gaze fixation to regions of interest and video item.

Reaction time of target selection (via mouse click).

### 5.3.7 Procedure

35 participants (aged 18-35, 24 = female, recruited via email announcements) took part in the study. Each participant saw all 160 items, but they were placed in four blocks of 40 with each composite image only appearing once in each block. The order in which the blocks were displayed was counterbalanced across participants. Each participant took approximately 30-40 minutes to complete the experiment. Eye gaze was recorded using a Tobii X120 remote desk-mounted eye tracker sampling at 60Hz. Arrays and videos were displayed on a Dell 17” flat panel monitor with a content area of 1280x1024 pixels. At the beginning of the experiment, calibration was conducted using a nine point calibration procedure. Following this, the participant read on screen instructions (see appendix) before completing eight practice trials. Participants were encouraged to ask questions throughout the practice. Each trial begins with a fixation cross in the centre of the screen which the participant had to click to start the trial. Following the click the array was displayed for 4000ms before the video appeared in the centre of the screen. The

onset of the word “draw” began at 5000ms and the participant was instructed to click on the target as soon as they have disambiguated it from the other items. The experiment automatically progressed to the next screen at 10,000ms. Self-paced breaks occurred at the end of every block.

### 5.3.8 Predictions

Comprehenders should be able to extract MANNER information regardless of whether it is communicated through speech or gesture. Based on this, it can be predicted that all conditions in which MANNER information is presented should result in earlier fixations than when MANNER information is not presented (speech -; gesture -). This means that the target should be fixated on proportionally more than the other items in the array prior to the GROUND DOP. Although Wharton (2009) does not suggest a comprehension procedure for gesture, it might be assumed that if gesture needs to be integrated as a natural sign, then participants will be slower to process speech with gesture (speech +; gesture +) than speech on its own (speech +; gesture -). In this case, gesture should not provide comprehender with a processing advantage. Therefore, the condition in which MANNER information is presented in both speech and gesture should not result in shorter latencies before target fixations than the condition in which MANNER is presented through only speech. When MANNER information is only presented through gesture, it should be slower than when MANNER is in speech but faster than when it is not presented. However, if speech and gesture are processed according to Enfield (2009b) then it should be expected that gestural information is integrated from the first moments. What’s more, MANNER information through gesture might be expected to facilitate processing when produced in addition to MANNER information in speech.

In terms of reaction time, following Wharton (2009) there should be no difference in terms of the length of time it takes participants to select the correct object. However, following Enfield (2009b), it should be expected that reaction times in the condition in which MANNER information appears in both speech and gesture should be shorter than when MANNER is presented through one modality alone.

# Chapter 6

## Analysis of Gesture in the Visual world

### 6.1 Introduction

This chapter presents the data analysis for the visual world experiment described in the previous chapter.

In doing so, it assesses the following hypotheses:

#### 1. Eye gaze fixation hypotheses

- (a) *hypothesis 1. Multimodal disambiguation hypothesis:* the presentation of information through either gesture or speech (conditions: speech +; gesture -, speech -, gesture +, or speech +; gesture +) will result in a significant increase in the proportion of looks to the target item earlier than when this information is not presented (condition speech -, gesture -).
- (b) *hypothesis 2. Speech advantage hypothesis:* the presentation of information through speech only (condition: speech +; gesture -) will result in an increase in the proportion of looks to the target item earlier than when it is presented through either speech or gesture (speech +; gesture + or speech -, gesture +).
- (c) *hypothesis 3. Composite disambiguation hypothesis:* the presentation of information through speech and gesture (condition: speech +; gesture +) will result in an increase in the proportion of looks to the target item earlier than

when it is presented through either speech or gesture (speech +; gesture - or speech -; gesture +).

- (d) *hypothesis 4. Composite disadvantage hypothesis*: the presentation of information through speech and gesture (condition: speech +; gesture +) will result in a decrease in the proportion of looks to the target item earlier than when it is presented through either speech or gesture (speech +; gesture - or speech -; gesture +).

## 2. Reaction time hypotheses

- (a) *hypothesis 5. Multimodal disambiguation hypothesis*: the presentation of information through either gesture or speech (conditions: speech +; gesture -, speech -; gesture +, or speech +; gesture +) will result in a reduction in reaction times in terms of target selection (via mouse click) compared to when it is presented through either speech or gesture (speech -; gesture -).
- (b) *hypothesis 6. Speech advantage hypothesis*: the presentation of information through speech only (condition: speech +; gesture -) will result in a reduction in reaction times in terms of target selection (via mouse click) compared to when it is presented through either speech or gesture (speech +; gesture + or speech -; gesture +).
- (c) *hypothesis 7. Composite disambiguation hypothesis*: the presentation of information through speech and gesture (condition: speech +; gesture +) will result in a reduction in reaction times in terms of target selection (via mouse click) than when it is presented through either speech or gesture (speech +; gesture - or speech -; gesture +).
- (d) *hypothesis 8. Composite disadvantage hypothesis*: the presentation of information through speech and gesture (condition: speech +; gesture +) will result in a reduction in reaction times in terms of target selection (via mouse click) than when it is presented through either speech or gesture (speech +; gesture - or speech -; gesture +).

The formulation of these hypotheses are grouped according to the principle that an early

increase in fixations to the target object is analogous to a decrease in reaction time. In relation to the different perspectives on gesture comprehension, hypotheses 1 and 5 should be true regardless of which theoretical position one adopts. They basically state that people will be able to extract information from speech or gesture, which would be predicted according to both the composite signal and natural sign perspectives. Hypotheses 2 and 6 are based on the assumption that speech provides an advantage not presented by gesture and that this advantage is only present when gesture is not. This view would be predicted in accordance to the natural sign perspective according to which speech and gesture are processed through different mechanisms. Hypotheses 3 and 7 are based on the idea that signals consisting of speech and gesture will provide an advantage over speech or gesture alone. In other words, there is an additive effect of modality. And, finally, hypotheses 4 and 8 are the opposite of 3 and 7 and relate to the same conditions. Hypotheses 4 and 8 are based on the idea that gesture will be tied to increased effort in terms of comprehension, therefore processing information across two modalities takes more effort than processing information across either modality alone. These two hypotheses would be predicted according to the natural sign perspective since the different mechanisms involved in comprehending different modalities. However, the opposite would be predicted by the composite signal view since the theory rests on the idea that the speech and gesture composites are produced opportunistically. Table 6.1 presents the relationship between the speech and gesture conditions and hypotheses.

The analysis is broken into three main sections. First, fixation data is explored through plots depicting the proportion of looks to each object in the array over time. Second, the data is analysed using linear mixed effects analyses, which focus on fixations to target, fixations to character and fixations to distractors. Third, the analysis of the reaction time measure is presented. Following the analysis, there is a discussion of how these results relate the predictions outlined in the previous chapter and to previous literature.

	+speech;+gesture	+speech;-gesture	-speech;+gesture	-speech;-gesture
+speech; +gesture		Hypotheses 2 & 6	Hypotheses 3 & 7 and Hypotheses 4 & 8	Hypotheses 1 & 5
+speech; - gesture	Hypotheses 3 & 7 and Hypotheses 4 & 8		Hypotheses 2 & 6	Hypotheses 1 & 5
-speech; +gesture	Hypotheses 3 & 7 and Hypotheses 4 & 8	Hypotheses 2 & 6		Hypotheses 1 & 5
-speech; - gesture	Hypotheses 1 & 5	Hypotheses 1 & 5	Hypotheses 1 & 5	

Table 6.1: Relationship between conditions and hypotheses

## 6.2 Data Preparation

Calibration was conducted at the beginning of every trial, which meant that degradation in the validity of eye tracking data was not consistent throughout. In order to overcome this fact, following the recommendations of the Tobii Studio user manual all data points with a validity greater than 2 were removed. In order to minimise the effect of this manipulation, only fixations were removed, without removing all data associated with a particular item. This meant that all 35 participants were included in the analysis but the amount of data each participant contributes to the study is determined by the validity of their data.

## 6.3 Data Exploration

The analysis below uses a measure of “target advantage”, taken from Kronmuller and Barr (2007), however, first it is useful to explore the data visually comparing the proportion (over 50ms time windows) of looks to each composite image in the array over time. Figure 6.1 provides four plots, one for each condition.

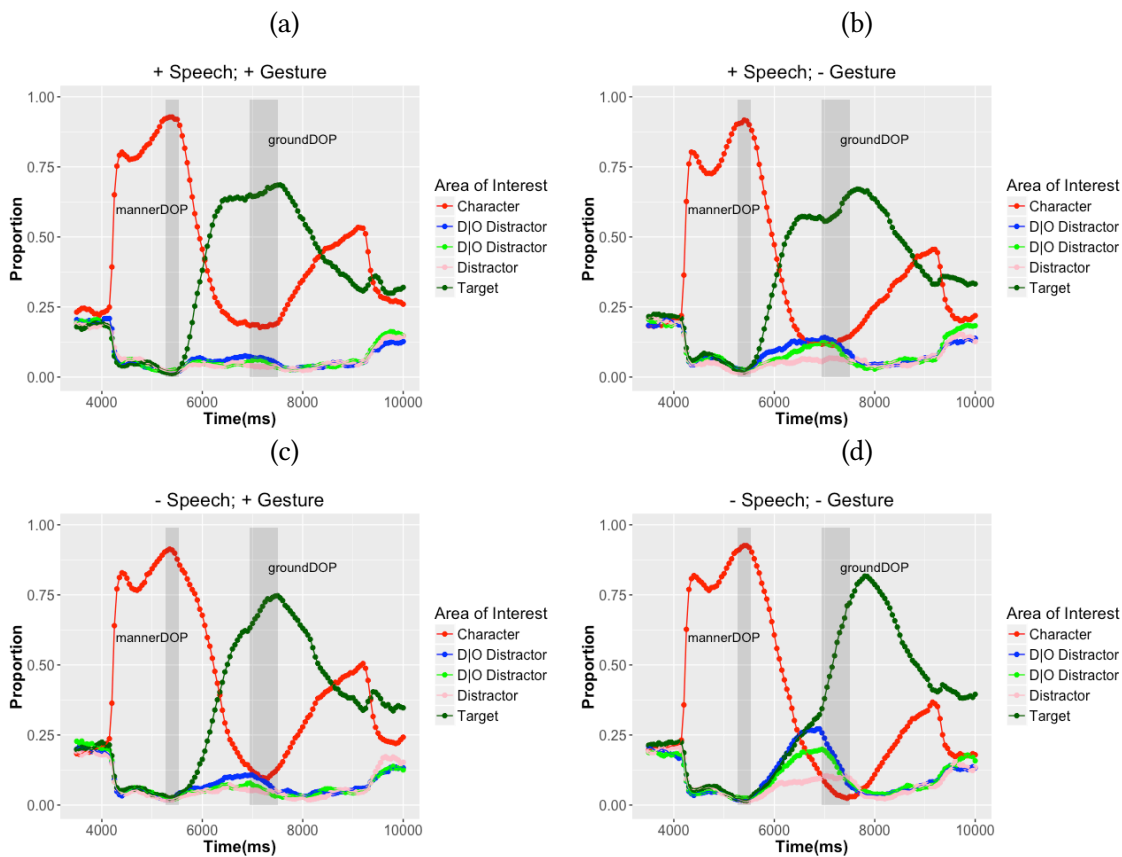


Figure 6.1: Plots showing proportion of looks to each item in the array by condition

The onset of the MANNER elements in speech and gesture and the offset of the MANNER element in speech are represented in figure 6.1 as the greyed out area labelled mannerPOD. The GROUND onset and offset are represented in the second greyed out area labelled groundPOD. The onset of the MANNER element occurs at 5270ms and the next word, “line”, starts at 5540ms, and therefore the earliest fixations to the target based on the participant disambiguating the target can be expected at around 5815ms (on the assumption that it takes 200ms to plan a saccade and 75ms to comprehend the word). The line representing looks to target shows an increase at this point in all conditions except the -speech; -gesture condition. This suggests that participants fixate on the target when MANNER information is given in speech and/or gesture. Furthermore, in this condition, the proportion of looks to target seems to deviate around 7000ms which is roughly the point at which the landmark is mentioned (i.e., the second POD).

Additionally, following the MANNERPOD, the drop in looks to character appears to be at a slower rate in the -speech; +gesture condition than in any other condition. This could mean that participants continue to fixate on the character for longer when gesture



is the only source of disambiguating information. Finally, there seems to be a different distribution of fixations to the competitors and distractors in the different conditions. In +speech; +gesture and the -speech; +gesture condition, fixations to all non-target items do not appear to increase following the MANNERPOD. However, in the +speech; -gesture and -speech; -gesture condition, there does seem to be prolonged fixations on non-target items. This finding is to be expected for the -speech; -gesture condition because it would be impossible for the participants to distinguish between the target item and DIRECTION & ORIENTATION competitors until the GROUNDPOD. This is exactly what seems to be reflected in the plots, with fixations to competitors dropping away after the onset of the GROUNDPOD. However, the prolonged looks to non-target items is harder to explain for the +speech; -gesture condition since participants are presented with disambiguating information during the MANNERPOD.

These observations regarding character fixations and competitor fixations will be taken up below. However, first this analysis focusses on target fixations.

## 6.4 Target Advantage

In order to further explore the observations of 6.1, a target advantage score was calculated. This was created by subtracting the average sum of fixations to the DIRECTION & ORIENTATION competitors and the distractor (see figure 5.6) from the average proportion of looks to target over 50ms time bins. This calculation gives a number between -1 and 1 where 1 means that the target is the only element that was fixated on and -1 means that only non-targets (competitors and the distractors) were focussed on competitors or the distractor were fixated on. Scores of 0 mean that fixations are split equally between the target and the non-targets. Character fixations were excluded from the analysis before average proportions were calculated. The scores over the 50ms timebins were then averaged over 300ms time windows based on the observations relating to figure 6.1 above and the different PODs. The nine time window are (in ms) 5250-5550, 5550-5850, 5850-6150, 6150-6450, 6450-6750, 6750-7050, 7050-7350, 7350-7650, 7650-7950. The third time-window depicts the first point at which fixations to the target can be expected. This represents a larger number of time windows than is standard in the eye tracking

literature. However, this seems warranted due to the differences in fixations between MANNERPOD and GROUNDPOD (shown in figure 6.1) and, as will be discussed below, target selections through mouse clicks were not made before 7650ms. Therefore, these time windows represent the time between earliest possible disambiguation (i.e., the onset of MANNERPOD) to the average point at which participants had selected the correct item.

Table 6.2 presents target advantage by condition in each of the nine time windows. This table is visually demonstrated in figure 6.4. The conditions are the same as those in figure 6.1 above.

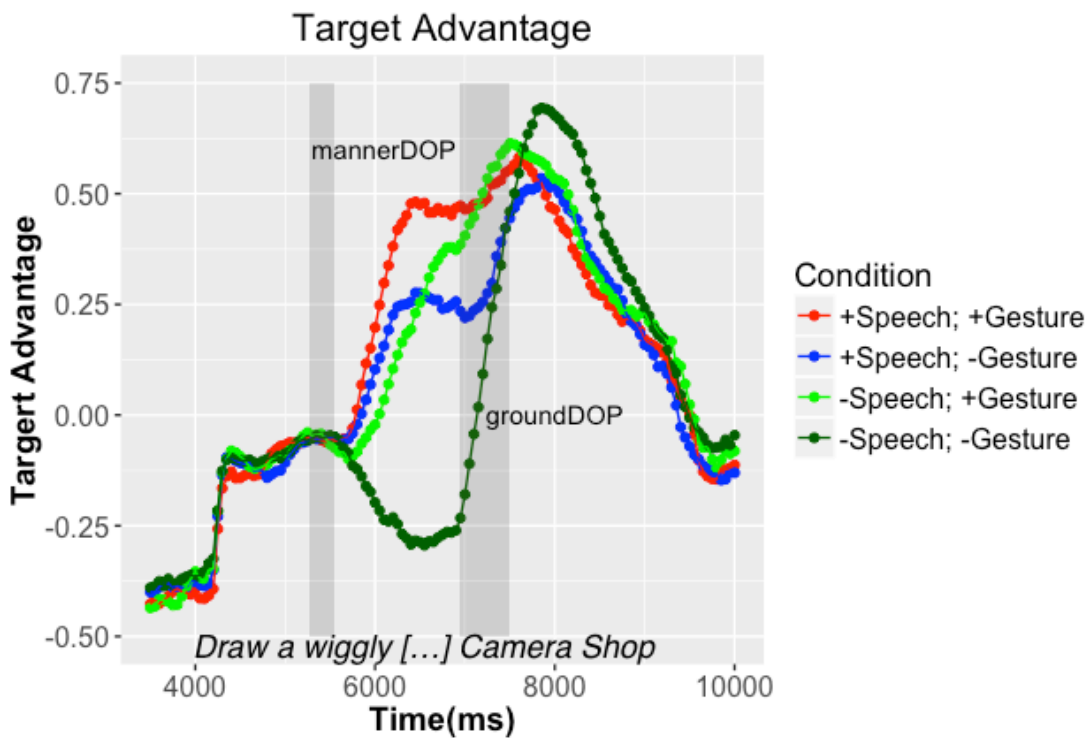


Figure 6.2: Target advantage against Time

---

Condition	Mean Target Advantage (SE)
<b>Window 1: 5250-5550ms</b>	
<i>-speech; -gesture</i>	-0.06 (0.01)
<i>+speech; -gesture</i>	-0.06 (0.02)
<i>-speech; +gesture</i>	-0.06 (0.01)

<i>+speech; +gesture</i>	-0.07 (0.02)
<b>Window 2: 5550-5850ms</b>	
<i>-speech; -gesture</i>	-10 (0.02)
<i>+speech; -gesture</i>	-0.07 (0.03)
<i>-speech; +gesture</i>	-0.09 (0.02)
<i>+speech; +gesture</i>	-0.06 (0.03)
<b>Window 3: 5850-6150ms</b>	
<i>-speech; -gesture</i>	-0.20 (0.03)
<i>+speech; -gesture</i>	0.08 (0.04)
<i>-speech; +gesture</i>	-0.02 (0.03)
<i>+speech; +gesture</i>	0.16 (0.05)
<b>Window 4: 6150-6450ms</b>	
<i>-speech; -gesture</i>	-0.28 (0.05)
<i>+speech; -gesture</i>	0.24 (0.05)
<i>-speech; +gesture</i>	0.13 (0.04)
<i>+speech; +gesture</i>	0.41 (0.05)
<b>Window 5: 6450-6750ms</b>	
<i>-speech; -gesture</i>	-0.30 (0.05)
<i>+speech; -gesture</i>	0.26 (0.05)
<i>-speech; +gesture</i>	0.29 (0.05)
<i>+speech; +gesture</i>	0.46 (0.06)
<b>Window 6: 6750-7050ms</b>	
<i>-speech; -gesture</i>	-0.24 (0.05)
<i>+speech; -gesture</i>	0.24 (0.05)
<i>-speech; +gesture</i>	0.37 (0.05)
<i>+speech; +gesture</i>	0.44 (0.06)
<b>Window 7: 7050-7350ms</b>	
<i>-speech; -gesture</i>	0.09 (0.05)
<i>+speech; -gesture</i>	0.28 (0.05)
<i>-speech; +gesture</i>	0.49 (0.05)

<i>+speech; +gesture</i>	0.48 (0.06)
<b>Window 8: 7350-7650ms</b>	
<i>-speech; -gesture</i>	0.44 (0.05)
<i>+speech; -gesture</i>	0.43 (0.04)
<i>-speech; +gesture</i>	0.58 (0.04)
<i>+speech; +gesture</i>	0.55 (0.05)
<b>Window 9: 7650-7950ms</b>	
<i>-speech; -gesture</i>	0.65 (0.04)
<i>+speech; -gesture</i>	0.50 (0.04)
<i>-speech; +gesture</i>	0.57 (0.04)
<i>+speech; +gesture</i>	0.52 (0.05)

Table 6.2: Target advantage by condition

### Model fitting procedure

The critical analysis involved linear mixed effects models being created for whether MANNER was present/absent in gesture and speech. These were created for each of the nine time windows specified above. For each model, target advantage was used as the outcome variable, participant and item were taken as random effects and condition (speech & gesture) were used as fixed effects. Additionally, random slopes for (speech and gesture) were included when they created a model that was a better fit for the data. In other words, the model fitting procedure began by attempting to fit the full model and simplifying the model if it overfitted the data. This was achieved by incrementally adding slope terms to the full model (without slopes) and comparing it to the full model with slopes. If the slopes produced a model which was a significantly better fit for the data then they were included in the analysis. All models included random slopes for the participant, but in the table below those analyses that do include a random slope term for the item are marked by a \*. This process is in line with the keep it maximal procedure of Barr et al. (2013). An outline of each analysis and the models used can be found in the

appendix.

Following the method of Winter (2013), significance values were generated using likelihood ratio tests, comparing the models to an intercept only model, using R's anova function. The intercept only model represents the mean target advantage for all conditions. Therefore, a significant result for a model suggests that it is significantly better fitted to the data than the mean. Interaction values were generated by comparing the full model (speech\*gesture) to a reduced model (speech + gesture). Paired contrasts were conducted post hoc in order to further explore the results.

For each of the time windows the following models were used:

#### **Intercept Only**

target advantage  $\sim 1 + (1 + \text{speech} + \text{gesture} \mid \text{Participant}) + (1 + \text{speech} + \text{gesture} \mid \text{Item})$

#### **Speech**

target advantage  $\sim \text{speech} + (1 + \text{speech} + \text{gesture} \mid \text{Participant}) + (1 + \text{speech} + \text{gesture} \mid \text{Item})$

#### **Gesture**

target advantage  $\sim \text{gesture} + (1 + \text{speech} + \text{gesture} \mid \text{Participant}) + (1 + \text{speech} + \text{gesture} \mid \text{Item})$

#### **Full (including the interaction of the predictor variables**

target advantage  $\sim \text{speech} * \text{gesture} + (1 + \text{speech} + \text{gesture} \mid \text{Participant}) + (1 + \text{speech} + \text{gesture} \mid \text{Item})$

#### **Reduced (including predictor variables and main effects**

target advantage  $\sim \text{speech} + \text{gesture} + (1 + \text{speech} + \text{gesture} \mid \text{Participant}) + (1 + \text{speech} + \text{gesture} \mid \text{Item})$

Table 6.4 presents the findings for the analyses for each of the time windows. The first two rows for each time window (“Speech” and “Gesture”) demonstrate whether or not presenting MANNER information in either of these modalities has a significant effect on target advantage. The third row (“Speech\*Gesture”) represents whether or not there is

an interaction between the effect of the two modalities. Significant findings are marked in bold with  $<0.05$  as the threshold. Additionally, any p-value  $>0.001$  is written out in full. The column labelled “Target Advantage (SE)” provides the estimate for the change in target advantage based on inclusion of MANNER information in speech and/or gesture. The  $\chi^2$ (DF) column shows the result of the comparison of the model to intercept only model. Lastly, the p-value is the p-value associated with that model comparison.

---

	Mean Target Advantage (SE)	$\chi^2$ (DF)	p-value
<b>Window 1: 5250-5550ms</b>			
<i>Speech</i>	-0.008 (0.016)	0.5115(10)	0.4745
<i>Gesture</i>	-0.002 (0.016)	0.0585(10)	0.809
<i>Speech*Gesture</i>	-0.0019 (0.019)	0.0103(12)	0.9192
<b>Window 2: 5550-5850ms</b>			
<i>Speech</i>	0.032 (0.02)	1.8052(10)	0.1791
<i>Gesture</i>	0.012 (0.03)	0.5642(10)	0.4526
<i>Speech*Gesture</i>	-0.0002 (0.03)	0(12)	0.9955
<b>Window 3: 5850-6150ms</b>			
<i>Speech</i>	0.28 (0.03)	16.478(10)	<b>&lt; 0.001</b>
<i>Gesture</i>	0.17 (0.04)	4.1528(10)	<b>0.04157</b>
<i>Speech*Gesture</i>	-0.09 (0.05)	3.0229(12)	0.0821
<b>Window 4: 6150-6450ms*</b>			
<i>Speech</i>	0.52 (0.05)	23.184(15)	<b>&lt;0.001</b>
<i>Gesture</i>	0.42 (0.06)	17.993(15)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	-0.24 (0.07)	11.175(17)	<b>&lt;0.001</b>
<b>Window 5: 6450-6750ms</b>			
<i>Speech</i>	0.56 (0.07)	23.028(10)	<b>&lt;0.001</b>
<i>Gesture</i>	0.58 (0.06)	34.624(10)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	-0.38 (0.08)	20.641(12)	<b>&lt;0.001</b>

**Window 6: 6750-7050ms**

<i>Speech</i>	0.47 (0.07)	13.418(10)	<b>0.02718</b>
<i>Gesture</i>	0.60 (0.07)	41.497(10)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	-0.39 (0.08)	21.244(12)	<b>&lt;0.001</b>

**Window 7: 7050-7350ms\***

<i>Speech</i>	0.19 (0.07)	1.1233(10)	0.2892
<i>Gesture</i>	0.40 (0.06)	31.893(10)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	-0.19 (0.08)	40.013	<b>0.01873</b>

**Window 8: 7350-7650ms**

<i>Speech</i>	-0.006 (0.053)	0.7896(10)	0.3742
<i>Gesture</i>	0.15 (0.051)	11.143(10)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	-0.032 (0.062)	0.2732(12)	0.6012

**Window 9: 7650-7950ms**

<i>Speech</i>	-0.15 (0.04)	5.3348(10)	<b>0.0209</b>
<i>Gesture</i>	-0.08 (0.04)	0.436(10)	0.5091
<i>Speech* Gesture</i>	0.104 (0.043)	5.8089(12)	<b>0.01594</b>

Table 6.4: Target Advantage Analysis

**Window 1: 5250-5550ms**

Window 1 does not reveal a significant effect of either speech or gesture on target advantage. This suggests that participants are not fixating on the target before the onset of the MANNER in either modality, nor during its production. The mean target advantage associated with each condition shows that all are related to a negative target advantage score, which means that there was a slight preference to fixate on non-target items.

**window 2: 5550 - 5850ms**

This window represents the point in between the offset of the MANNER element but prior to the point one might expect fixations to the target (based on the 275ms lag to pro-

gram a saccade). Once again, the mean target advantage associated with each window is marginally negative.

### window 3: 5850-6150ms

This window represents the point at which the first stimulus driven fixations on the target might be expected. It is demonstrated that MANNER presented in both speech ( $\chi^2(10) = 16.478, p = < 0.001$ ) and gesture ( $\chi^2(10) = 4.528, p = 0.04157$ ) result in an increased target advantage. Speech increases target advantage by  $0.28 \pm 0.03$  (SE) and gesture increased it by  $0.17 \pm 0.04$  (SE). Paired contrasts are presented in table 6.5:

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.27614118	0.04857856	-5.684	<.0001
-speech;-gesture - -speech;+gesture	-0.17365134	0.04301702	-4.037	0.0003
+speech;-gesture - -speech;+gesture	0.10248984	0.04779385	2.144	0.1392
-speech;-gesture - +speech;+gesture	-0.35946888	0.05852432	-6.142	<.0001
+speech;-gesture - +speech;+gesture	-0.08332769	0.04299699	-1.938	0.2121
-speech;+gesture - +speech;+gesture	-0.18581754	0.04820921	-3.854	0.0007

Table 6.5: Contrasts for Time Window 3

Table 6.5 shows that MANNER information presented through speech or gesture increases target advantage compared to the condition where MANNER information is not presented. Speech seems to be consistently related to an increase in target advantage this is not the case for gesture. Compared to the condition in which speech conveys MANNER information, the addition of gesture does increase target advantage. However, compared to the condition in which only gesture conveys MANNER information the addition of speech does significantly increase target advantage. Comparing these results to table 6.2, gesture on its own (-speech;+gesture) is still associated with a negative mean target advantage. This suggests that the main effect of gesture is related to the condition in which both speech and gesture convey MANNER (+speech:+gesture). These findings suggest that participants gestures are having an effect compared to when no MANNER information is conveyed, but speech is clearly influencing people to a greater extent early



on.

**window 4: 6150-6450ms**

Both speech ( $\chi^2(15) = 23.184, p = < 0.001$ ) and gesture ( $\chi^2(15) = 17.993, p = < 0.001$ ) significantly affect target advantage. The presence of MANNER in speech results in an increase of  $0.52 \pm 0.05$  (SE) and the presence of MANNER in gesture results in an increase of  $0.42 \pm 0.06$ . Paired contrasts are shown in table 6.6.

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.5192151	0.06756722	-7.684	<.0001
-speech;-gesture - -speech;+gesture	-0.4166329	0.05661797	-7.359	<.0001
+speech;-gesture - -speech;+gesture	0.1025822	0.06132201	1.673	0.3381
-speech;-gesture - +speech;+gesture	-0.6923410	0.06672331	-10.376	<.0001
+speech;-gesture - +speech;+gesture	-0.1731259	0.05295233	-3.269	0.0059
-speech;+gesture - +speech;+gesture	-0.2757082	0.05224644	-5.277	<.0001

Table 6.6: Contrasts for Time Window 4

Table 6.6 shows that MANNER in both speech and gesture increase target advantage compared to when this information is not presented. It is clear that participants were able to extract information from speech and gesture. Adding speech to gesture increases target advantage, but so does gesture when added to speech. This suggests that when taken together, information conveyed through speech and gesture facilitate the comprehension of utterances, compared to when speech and gesture convey information separately. Linking this to 6.2, gesture presented on its own is now associated with a positive mean target advantage score ( $M = 0.13, SE = 0.04$ ). However, speech presented on its own is associated with a higher mean target advantage ( $M = 0.24, SE = 0.05$ ) and the mean for speech and gesture is higher still ( $M = 0.41, SE = 0.05$ ). This window suggests that the facilitatory effect of the presentation of MANNER information goes from speech and gesture, speech alone, and then gesture alone. However, it is important to highlight that the contrasts revealed that gesture alone and speech alone are not significantly different. So speech and gesture are equally good at disambiguating the target, but both together are better.

**window 5: 6450-6750**

Both speech ( $\chi^2(10) = 23.028, p = < 0.001$ ) and gesture ( $\chi^2(10) = , p = < 0.001$ ) significantly affect target advantage. Speech results in an increase of  $0.56 \pm 0.07$  (SE) and gesture results in an increase of  $0.58 \pm 0.06$  (SE). This window, therefore, represents the first point at which gesture is having a greater effect on target advantage than speech. Paired contrasts are shown in table 6.7.

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.55887404	0.06593304	-8.476	<.0001
-speech;-gesture - -speech;+gesture	-0.58068076	0.06261544	-9.274	<.0001
+speech;-gesture - -speech;+gesture	-0.02180672	0.07049708	-0.309	0.9897
-speech;-gesture - +speech;+gesture	-0.75939895	0.07022140	-10.814	<.0001
+speech;-gesture - +speech;+gesture	-0.20052491	0.06262884	-3.202	0.0075
-speech;+gesture - +speech;+gesture	-0.17871819	0.06529412	-2.737	0.0315

Table 6.7: Contrasts for Time Window 5

Once again speech and gesture result in an significantly increased target advantage. There is no difference between the either speech on its own or gesture on its own. Furthermore, speech added to gesture still provides a significant advantage as does gesture added to speech.

**window 6: 6750 - 7050**

Speech ( $\chi^2(10) = 13.417, p = 0.02718$ ) and gesture ( $\chi^2(10) = 41,497, p = < 0.001$ ) both significantly affect target advantage. For this time window, speech increases target advantage by  $0.47 \pm 0.07$  (SE) and gesture increases it by  $0.60 \pm 0.07$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.47269040	0.06739036	-7.014	<.0001
-speech;-gesture - -speech;+gesture	-0.60222261	0.06451060	-9.335	<.0001
+speech;-gesture - -speech;+gesture	-0.12953222	0.07211913	-1.796	0.2751
-speech;-gesture - +speech;+gesture	-0.68008794	0.07257312	-9.371	<.0001
+speech;-gesture - +speech;+gesture	-0.20739754	0.06454453	-3.213	0.0072
-speech;+gesture - +speech;+gesture	-0.07786533	0.06674212	-1.167	0.6480

Table 6.8: Contrasts for Time Window 6

The contrasts in table 6.8 demonstrate that speech and gesture are still resulting in an increase in target advantage. However, speech is no longer presenting an advantage when added to gesture (based on the contrast between -speech;+gesture and +speech;+gesture). Adding gesture to speech still provides a benefit. The mean target advantage scores for this window demonstrate that these findings are not the result of a drop in drop in target advantage for the +speech;-gesture condition ( $M = 0.24$ ,  $SE = 0.05$ ), but are due to a continued increase in the -speech;+gesture condition ( $M = 0.37$ ,  $SE = 0.05$ ). Therefore, this window suggests that while there was an early advantage of speech as shown in window 3, gesture's effect is slower but stronger. In other words, it seems that participants fixate on the target more when the information is processed in gesture than in speech. Importantly, this also carries over to the +speech;+gesture condition which has the highest mean target advantage score ( $M = 0.44$ ,  $SE = 0.06$ ). This 'late effect' of gesture is an unexpected finding.

#### window 7: 7050-7350

This window represents the first window where early fixations as a result of the GROUNDPOD might be expected. Here, for the first time since the MANNERPOD, speech is not significant ( $\chi^2(10) = 1.1233$ ,  $p = 0.2892$ ). Gesture is still significant ( $\chi^2(10) = 31.893$ ,  $p = < 0.001$ ). However, speech is still tied to an increase in target advantage of  $0.19 \pm 0.07$  (SE), but gesture's affect is greater, with gesture increasing target advantage by  $0.40 \pm 0.06$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.185269248	0.06760151	-2.741	0.0312
-speech;-gesture - -speech;+gesture	-0.395722797	0.06303360	-6.278	<.0001
+speech;-gesture - -speech;+gesture	-0.210453550	0.06957160	-3.025	0.0133
-speech;-gesture - +speech;+gesture	-0.387948037	0.07443702	-5.212	<.0001
+speech;-gesture - +speech;+gesture	-0.202678789	0.06305579	-3.214	0.0072
-speech;+gesture - +speech;+gesture	0.007774761	0.06699704	0.116	0.9994

Table 6.9: Contrasts for Time Window 7

Table 6.9 shows that speech and gesture are still significantly resulting in a higher target advantage than the condition in which MANNER is not presented. Speech is still not increasing target advantage in addition to gesture. However, gesture is still resulting in an increase in addition to speech. Also, in this window gesture on its own (-speech;+gesture) is significantly different from speech on its own (+speech;-gesture), suggesting that there is a significant difference between the effect of speech and gesture when they occur alone. This difference demonstrates that gesture (without speech) results in an increase in target advantage, but speech (without gesture) does not. These findings are, once again, reflected in the mean target advantage scores associated with this time window. While the mean target advantage for the +speech;-gesture is similar to time window 6 ( $M = 0.28$ ,  $SE = 0.05$ ), -speech;+gesture has continued to rise ( $M = 0.49$ ,  $SE = 0.05$ ). These findings show that these results are not the product of a reduction in the effect of speech but a further increase in the effect of gesture.

#### window 8: 7350-7650

This window continues the trend of window 7. Speech is still not significantly affecting target advantage ( $\chi^2(10) = 0.7896$ ,  $p = 0.3742$ ), while gesture does ( $\chi^2(10) = 11.143$ ,  $p = < 0.001$ ). Furthermore, for the first time, although it is small, speech is actually accounting for a reduction in target advantage of  $-0.006 \pm 0.05$ . Gesture increases target advantage, resulting in an increase of  $0.15 \pm 0.05$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	0.005554458	0.05307316	0.105	0.9996
-speech;-gesture - -speech;+gesture	-0.149592406	0.05063115	-2.955	0.0165
+speech;-gesture - -speech;+gesture	-0.155146864	0.05493534	-2.824	0.0245
-speech;-gesture - +speech;+gesture	-0.111711966	0.06225248	-1.794	0.2759
+speech;-gesture - +speech;+gesture	-0.117266423	0.05064559	-2.315	0.0945
-speech;+gesture - +speech;+gesture	0.037880441	0.05263275	0.720	0.8893

Table 6.10: Contrasts for Time Window 8

Table 6.10 demonstrates that gesture is still significantly affecting target advantage. This window suggests that gesture is still resulting in an increased target advantage score, but speech is not, nor is speech and gesture when they occur together. This and window 7 demonstrate a drop in the effect of speech independent of the effect of gesture. However, the mean target advantage scores once again reveal that these findings are not because participants are not looking at the target in the speech only condition, but because fixations in that condition are regressing to the mean. One of the reasons for this is that this window represents the point in which participants in the -speech;-gesture condition are fixating on the target, represented by the heightened mean target advantage associated with the -speech;-gesture condition ( $M = 0.44$ ,  $SE = 0.05$ ).

### **window 9: 7700-8000**

This final window continues the trend of window 8, however speech is now having a significant affect on target advantage ( $\chi^2(10) = 5.3348$ ,  $p = 0.0209$ ) whereas gesture is not ( $\chi^2(10) = 0.436$ ,  $p = 0.5091$ ). However, for this window, both speech and gesture account for a reduction in target advantage. Speech results in reduction of  $-0.15 \pm 0.04$  (SE) and gesture results in a reduction of  $-0.08 \pm 0.04$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	0.15008872	0.04381988	3.425	0.0034
-speech;-gesture - -speech;+gesture	0.08432947	0.03644843	2.314	0.0949
+speech;-gesture - -speech;+gesture	-0.06575925	0.04449435	-1.478	0.4509
-speech;-gesture - +speech;+gesture	0.13046504	0.05167420	2.525	0.0562
+speech;-gesture - +speech;+gesture	-0.01962369	0.03644799	-0.538	0.9497
-speech;+gesture - +speech;+gesture	0.04613556	0.04357399	1.059	0.7146

Table 6.11: Contrasts for Time Window 9

Table 6.11 reveals that the results for this window are in the opposite direction to those for the previous time windows. It is the case that in those conditions in which MANNER information is not present that target advantage is highest. The reason for this is not because participants are not fixating on the target in the +speech and +gesture conditions, but because they are fixating more in its absence (i.e., in the -speech;-gesture condition). This suggests that in this window the mean target advantage associated with the -speech;-gesture condition is higher than all other conditions ( $M = 0.65$ ,  $SE = 0.04$ ). It is likely that the reason for this finding is due to the fact that in the +MANNER conditions, participants have already selected the target object and are beginning to focus on the character again.

### Summary of Results for Target Advantage

There are three general findings in these results. First, MANNER information can be extracted from both speech and gesture. This is in line with hypothesis 1, which stated that the presentation of MANNER information through speech or gesture will result in an earlier increase in the proportion of looks to the target item and that hypothesis 2 is not entirely accurate, speech and gesture both result in this effect. The second finding is that there is an early advantage of speech over gesture (shown in window 3). This seems to suggest an advantage of speech over gesture, which is partially in line with hypothesis 2. However, it is also shown that when speech and gesture both present MANNER information, participants fixate on the target sooner than when only speech presents it (although this was not significant). This is important because it suggests that

the occurrence of gesture alongside speech is not resulting in a decrease in looks to target. In other words, this favours hypothesis 3 and goes against hypothesis 4. Taking these findings together, it seems that participants are more likely to fixate on the target item earlier when both speech and gesture or just speech disambiguates it than when only gesture does. However, gesture results in a greater number of fixations than when MANNER is not presented at all.

The third finding, which is somewhat unexpected, is that there is a ‘late’ effect of gesture. As the stimulus utterances unfold, target advantage gradually increases in all conditions that include MANNER. However, from time window 6 onwards, the increase in those conditions where MANNER was depicted in gesture, target advantage continues to rise, but in the condition in which it is only presented in speech (+speech;-gesture) this is not the case. This suggests that while there was an early advantage of speech, participants potentially began looking at other areas of the display as the trial continued. This finding suggests that participants are surer of what is being referred to in the conditions where that information is presented through speech but not gesture.

In order to further explore findings 2 and 3, it is important to explore where participants were looking when they were (or were not) fixating on the target. In order to explore this, the next section will focus on character advantage, competitor advantage, and response time in terms of target selection.

## 6.5 Character Advantage

Character advantage is calculated in the same manner as target advantage, however, fixations on the distractor or the competitors are subtracted from fixations on the character. Fixations to target were not included in the calculation. Figure 6.3 shows a visual representation of target advantage.

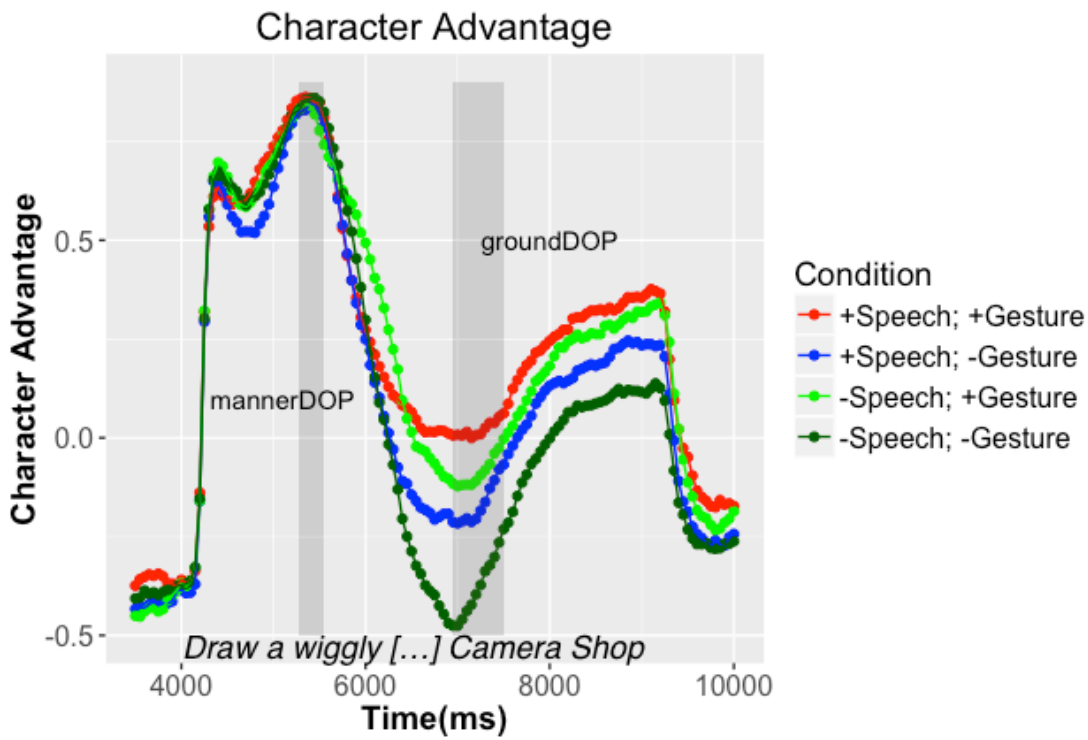


Figure 6.3: Character Advantage against Time

The following table shows the mean character advantage for each condition across the nine time windows.

Condition	Mean Character Advantage (SE)
<b>Window 1: 5250-5550ms</b>	
<i>-speech; -gesture</i>	0.84 (0.03)
<i>+speech; -gesture</i>	0.80 (0.03)
<i>-speech; +gesture</i>	0.80 (0.04)
<i>+speech; +gesture</i>	0.81 (0.04)
<b>Window 2: 5550-5850ms</b>	
<i>-speech; -gesture</i>	0.66 (0.05)
<i>+speech; -gesture</i>	0.55 (0.06)
<i>-speech; +gesture</i>	0.64 (0.05)
<i>+speech; +gesture</i>	0.54 (0.06)
<b>Window 3: 5850-6150ms</b>	



<i>-speech; -gesture</i>	0.30 (0.06)
<i>+speech; -gesture</i>	0.21 (0.06)
<i>-speech; +gesture</i>	0.46 (0.06)
<i>+speech; +gesture</i>	0.23 (0.06)

**Window 4: 6150-6450ms**

<i>-speech; -gesture</i>	-0.08 (0.06)
<i>+speech; -gesture</i>	-0.04 (0.05)
<i>-speech; +gesture</i>	0.21 (0.06)
<i>+speech; +gesture</i>	0.10 (0.05)

**Window 5: 6450-6750ms**

<i>-speech; -gesture</i>	-0.35 (0.05)
<i>+speech; -gesture</i>	-0.19 (0.04)
<i>-speech; +gesture</i>	-0.03 (0.05)
<i>+speech; +gesture</i>	-0.02 (0.04)

**Window 6: 6750-7050ms**

<i>-speech; -gesture</i>	-0.47 (0.04)
<i>+speech; -gesture</i>	-0.23 (0.04)
<i>-speech; +gesture</i>	-0.12 (0.04)
<i>+speech; +gesture</i>	-0.02 (0.04)

**Window 7: 7050-7350ms**

<i>-speech; -gesture</i>	-0.40 (0.03)
<i>+speech; -gesture</i>	-0.19 (0.04)
<i>-speech; +gesture</i>	-0.12 (0.03)
<i>+speech; +gesture</i>	-0.004 (0.04)

**Window 8: 7350-7650ms**

<i>-speech; -gesture</i>	-0.25 (0.03)
<i>+speech; -gesture</i>	-0.08 (0.04)
<i>-speech; +gesture</i>	-0.02 (0.03)
<i>+speech; +gesture</i>	0.07 (0.05)

**Window 9: 7650-7950ms**

<i>-speech; -gesture</i>	-0.08 (0.03)
<i>+speech; -gesture</i>	0.04 (0.04)
<i>-speech; +gesture</i>	0.10 (0.03)
<i>+speech; +gesture</i>	0.18 (0.04)

The following table shows the model comparisons for each time window. Models are identical to the ones used for target advantage with character advantage as the outcome variable.

---

	Character Advantage (SE)	$\chi^2$ (DF)	p-value
<b>Window 1: 5250-5550ms</b>			
<i>Speech</i>	-0.032 (0.016)	0.0675(10)	0.7951
<i>Gesture</i>	-0.04 (0.016)	0.587(10)	0.4436
<i>Speech*Gesture</i>	0.046 (0.019)	1.8237(12)	0.1769
<b>Window 2: 5550-5850ms</b>			
<i>Speech</i>	-0.11 (0.05)	7.3273(10)	<b>0.006791</b>
<i>Gesture</i>	0.019 (0.04)	0.1773(10)	0.6737
<i>Speech*Gesture</i>	0.009 (0.06)	0.0239(12)	0.8772
<b>Window 3: 5850-6150ms</b>			
<i>Speech</i>	-0.09 (0.06)	12.589(10)	< <b>0.001</b>
<i>Gesture</i>	0.16 (0.05)	5.4501(10)	<b>0.0197</b>
<i>Speech*Gesture</i>	-0.14 (0.07)	4.4752(12)	<b>0.03439</b>
<b>Window 4: 6150-6450ms</b>			
<i>Speech</i>	0.045 (0.05)	1.9374(10)	0.1639
<i>Gesture</i>	0.29 (0.05)	27.053(10)	< <b>0.001</b>
<i>Speech*Gesture</i>	-0.15 (0.06)	6.6621(12)	<b>0.009849</b>
<b>Window 5: 6450-6750ms</b>			

<i>Speech</i>	0.16 (0.05)	1.0224(10)	0.312
<i>Gesture</i>	0.32 (0.04)	45.5(10)	< <b>0.001</b>
<i>Speech*Gesture</i>	-0.11 (0.05)	4.742(12)	<b>0.02944</b>
<b>Window 6: 6750-7050ms</b>			
<i>Speech</i>	0.24 (0.04)	15.703(10)	< <b>0.001</b>
<i>Gesture</i>	0.34 (0.04)	34.939(10)	< <b>0.001</b>
<i>Speech*Gesture</i>	-0.13 (0.05)	6.583(12)	<b>0.0103</b>
<b>Window 7: 7050-7350ms</b>			
<i>Speech</i>	0.20 (0.04)	13.546(10)	< <b>0.001</b>
<i>Gesture</i>	0.28 (0.04)	40.067(10)	< <b>0.001</b>
<i>Speech*Gesture</i>	-0.09 (0.05)	3.4128(12)	<b>0.06469</b>
<b>Window 8: 7350-7650ms</b>			
<i>Speech</i>	0.17 (0.04)	4.6775(10)	<b>0.03056</b>
<i>Gesture</i>	0.23 (0.03)	27.001(10)	< <b>0.001</b>
<i>Speech*Gesture</i>	-0.09 (0.04)	4.4644(12)	<b>0.03461</b>
<b>Window 9: 7650-7950ms</b>			
<i>Speech</i>	0.13 (0.04)	2.8818(10)	0.08958
<i>Gesture</i>	0.18 (0.03)	26.585(10)	< <b>0.001</b>
<i>Speech*Gesture</i>	-0.05 (0.04)	1.6694(12)	0.1963

**window 1: 5250-5550ms**

Window 1 does not reveal a significant effect of either speech or gesture. This finding is because all conditions are associated with high character advantage scores, and therefore, participants are equally likely to be looking at the character in all conditions.

**window 2: 5550 - 5850ms**

In this window, speech significantly affects character advantage ( $\chi^2(10) = 7.3273, p = < 0.006791$ ) whereas gesture does not ( $\chi^2(10) = 0.1773, p = < 0.6737$ ). Further, speech

accounts for a reduction in character advantage of  $-0.11 \pm 0.05(\text{SE})$ , whereas gesture results in a minor increase  $0.019 \pm 0.04$ .

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	0.105989491	0.04736746	2.238	0.1132
-speech;-gesture - -speech;+gesture	0.018658149	0.04388896	0.425	0.9742
+speech;-gesture - -speech;+gesture	-0.087331342	0.04958735	-1.761	0.2922
-speech;-gesture - +speech;+gesture	0.115662997	0.04960519	2.332	0.0909
-speech;+gesture - +speech;+gesture	0.009673505	0.04390541	0.220	0.9962
+speech;-gesture - +speech;+gesture	0.097004848	0.04693194	2.067	0.1640

Table 6.14: Character Advantage contrasts for Time Window 2

Table 6.14 does not demonstrate any significant individual contrasts. Looking at the mean character advantage scores associated with this window demonstrates that those conditions in which MANNER is presented through speech, are associated with the lowest mean character advantage. Therefore the significant effect found in the model comparison is likely to be the accumulative effect of the two conditions in which MANNER information is presented in speech. This time window therefore represents the other side of early effect of MANNER presented in speech. Participants very quickly start to reduce fixations from the character, but before they start to fixate on the target. However, there is no difference between the condition in which MANNER is presented in gesture (-speech;+gesture) and the condition in which MANNER is not presented at all (-speech;-gesture). Therefore, this suggests that this early effect is the result of speech only. What’s more, gesture does not negatively affect the processing of speech in the condition where MANNER information is presented in both.

**window 3: 5850-6150ms**

This window is particularly interesting because it demonstrates speech and gesture are affecting character advantage in different directions. Both speech ( $\chi^2(15) = 23.184, p = < 0.001$ ) and gesture ( $\chi^2(15) = 23.184, p = < 0.001$ ) are significant. Speech is reducing character advantage by  $0.09 \pm 0.05$  (SE), whereas gesture is increasing it by  $0.16 \pm 0.05$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	0.09259093	0.05721570	1.618	0.3682
-speech;-gesture - -speech;+gesture	-0.16425698	0.05386118	-3.050	0.0123
+speech;-gesture - -speech;+gesture	-0.25684790	0.06018790	-4.267	0.0001
-speech;-gesture - +speech;+gesture	0.06985932	0.06522690	1.071	0.7072
-speech;+gesture - +speech;+gesture	-0.02273161	0.05384200	-0.422	0.9747
+speech;-gesture - +speech;+gesture	0.23411630	0.05670633	4.129	0.0002

Table 6.15: Character Advantage contrasts for Time Window 3

Table 6.15 shows that both speech on its own and gesture on its own are significantly different from the no MANNER condition. However, as suggested above these effects are in different directions. Gesture is responsible for a greater proportion of fixations to character and speech is responsible for a reduction in proportions to character. What’s more this is true when both speech and gesture convey MANNER information. What this suggests is that when crucial information is presented through speech then participants begin to fixate on the target. However, when crucial information is only presented through gesture, then participants continue to fixate on the character. One possible reason for this is that participants anticipate that further crucial information is going to be presented through gesture and therefore they need to continue to pay attention to it. This time window helps explain the early effect of speech shown before. In the condition where MANNER information is not presented, participants are beginning to stop fixating on the character, but they are not fixating on the target either. However, when MANNER information is presented only in gesture, participants continue to fixate on the character. These findings suggests that participants are able to distinguish crucial from not crucial gestured information and thus it suggests that comprehension of gesture occurs as early as the comprehension of speech.

**window 4: 6150-6450ms**

In time window 4, gesture significantly affects character advantage ( $\chi^2(10) = 27.053, p < 0.001$ ) whereas speech does not ( $\chi^2(10) = 1.9374, p = 0.1639$ ). However, both result in an increase in character advantage, with speech increasing it by  $0.045 \pm 0.05$  and gesture

increasing it by  $0.29 \pm 0.05$ .

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.04485297	0.05078938	-0.883	0.8136
-speech;-gesture - -speech;+gesture	-0.28519831	0.04794996	-5.948	<.0001
+speech;-gesture - -speech;+gesture	-0.24034534	0.05223288	-4.601	<.0001
-speech;-gesture - +speech;+gesture	-0.17958655	0.06049480	-2.969	0.0158
-speech;+gesture - +speech;+gesture	-0.13473358	0.04795530	-2.810	0.0256
+speech;-gesture - +speech;+gesture	0.10561176	0.05036159	2.097	0.1540

Table 6.16: Character Advantage contrasts for Time Window 4

Table 6.16 demonstrates that gesture (-speech;+gesture) increases target advantage compared to the no MANNER condition (-speech;-gesture) and when MANNER is only conveyed through speech (+speech;-gesture). Gesture on its own is also significantly different than gesture and speech together (+speech;+gesture). However, speech (+speech;-gesture) is not significantly different from speech and gesture or the no MANNER condition. However, speech and gesture presented together is significantly different from no MANNER. These findings show that crucial information presented through gesture leads to prolonged looks to the character. However, only when gesture is the only channel that presents that information is there a significant difference between speech and gesture. However, the average character advantages show that the condition in which MANNER information is present through speech and gesture is almost exactly halfway between the gesture only and speech only conditions.

These findings continue the trend of window 3. Interestingly, in this window, although the +speech;+gesture condition is associated with a higher character advantage than the +speech;-gesture condition, it is also associated with a higher target advantage. Therefore, this suggests that in the +speech;-gesture condition, participants are fixating on areas of the display that are not the target nor the character. This is also the case for the no MANNER condition.

**window 5: 6450-6750**

This window mirrors the findings of the last window. Gesture significantly affects character advantage ( $\chi^2(10) = 45.5, p = < 0.001$ ) whereas speech does not ( $\chi^2(10) = 1.0224, p = 0.312$ ). However, once again, both result in an increase in character advantage, with speech increasing it by  $0.16 \pm 0.05$  and gesture increasing it by  $0.32 \pm 0.04$ .

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.16016614	0.04489814	-3.567	0.0020
-speech;-gesture - -speech;+gesture	-0.32061667	0.04084325	-7.850	<.0001
+speech;-gesture - -speech;+gesture	-0.16045053	0.04604416	-3.485	0.0028
-speech;-gesture - +speech;+gesture	-0.36857581	0.05099835	-7.227	<.0001
+speech;-gesture - +speech;+gesture	-0.20840967	0.04085801	-5.101	<.0001
-speech;+gesture - +speech;+gesture	-0.04795914	0.04452589	-1.077	0.7035

Table 6.17: Character Advantage contrasts for Time Window 5

Table 6.17 shows that almost all contrasts are significant in time window 5. Crucially, there is not a significant difference between the -speech;+gesture condition and the +speech;+gesture condition. Participants are looking at the character equally in the conditions that include MANNER information in gesture. Interestingly, like window 4, this window the +speech;+gesture condition is associated with the highest target advantage. This suggests that while participants are looking at the character less when MANNER is only presented through speech, this is not because they are fixating on the target more. This would suggest that participants are fixating on the distractor or competitors.

**window 6: 6750 - 7050**

In this window both speech ( $\chi^2(10) = 15.703, p = < 0.001$ ) and gesture ( $\chi^2(10) = 34.030, p = < 0.001$ ) significantly affect character advantage. Speech accounts for an increase of  $0.24 \pm 0.04$  (SE), whereas gesture accounts for an increase of  $0.34 \pm 0.04$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.2404932	0.03935666	-6.111	<.0001
-speech;-gesture - -speech;+gesture	-0.3441294	0.04222057	-8.151	<.0001
+speech;-gesture - -speech;+gesture	-0.1036362	0.04460096	-2.324	0.0927
-speech;-gesture - +speech;+gesture	-0.4527428	0.04535124	-9.983	<.0001
+speech;-gesture - +speech;+gesture	-0.2122496	0.04224861	-5.024	<.0001
-speech;+gesture - +speech;+gesture	-0.1086134	0.03894116	-2.789	0.0271

Table 6.18: Character Advantage contrasts for Time Window 6

Table 6.18 reveals that almost every comparison is significant. Importantly, in this time window the condition in which MANNER information is presented in only gesture is actually resulting in a lower character advantage than when it is conveyed through both speech and gesture. This window is also the point at which the late effect of gesture was observed for target advantage. Therefore, this suggests that the low character advantage associated with gesture is different from the low character advantage associated with the speech only condition. The difference is that while gesture is increasing looks to the target, speech is doing this to a lesser extent. Speech, then, must be resulting in an increased number of looks to either the competitors or the distractor.

#### window 7: 7050-7350

This window, once again, is almost identical to the last one. Speech ( $\chi^2(10) = 13.546, p = < 0.001$ ) and gesture ( $\chi^2(10) = 40.067, p = < 0.001$ ) significantly affect character advantage. Speech increases character advantage by  $20 \pm 0.04$  (SE), whereas gesture increases it by  $28 \pm 0.04$  (SE).



Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.20301402	0.03912846	-5.188	<.0001
-speech;-gesture - -speech;+gesture	-0.27978613	0.03783450	-7.395	<.0001
+speech;-gesture - -speech;+gesture	-0.07677211	0.04206632	-1.825	0.2615
-speech;-gesture - +speech;+gesture	-0.39234107	0.04190584	-9.362	<.0001
+speech;-gesture - +speech;+gesture	-0.18932705	0.03785138	-5.002	<.0001
-speech;+gesture - +speech;+gesture	-0.11255494	0.03875800	-2.904	0.0193

Table 6.19: Character Advantage contrasts for Time Window 7

Table 6.19 reveals that the contrasts are in the same relationship as time window 6. The mean scores for each condition show that the conditions are in the same order as the window 6.

**window 8: 7350-7650**

This window is the same as the previous window, however the effect of speech has been reduced ( $\chi^2(10) = 4.6775, p = 0.03056$ ). Gesture is still significant ( $\chi^2(10) = 27.001, p < 0.001$ ). Speech results in an increase of  $17 \pm 0.04$ , whereas gesture results in an increase of  $23 \pm 0.03$ .

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.17431074	0.03687593	-4.727	<.0001
-speech;-gesture - -speech;+gesture	-0.23093512	0.03409276	-6.774	<.0001
+speech;-gesture - -speech;+gesture	-0.05662438	0.03765933	-1.504	0.4353
-speech;-gesture - +speech;+gesture	-0.31668914	0.04328775	-7.316	<.0001
+speech;-gesture - +speech;+gesture	-0.14237840	0.03410315	-4.175	0.0002
-speech;+gesture - +speech;+gesture	-0.08575403	0.03658900	-2.344	0.0883

Table 6.20: Character Advantage contrasts for Time Window 8

Table 6.20 continues the trend of the time window 7. However, this time gesture on its own is not significantly different from speech and gesture.

**window 9: 7700-8000**

In the final window speech is no longer significant ( $\chi^2(10) = 2.8818, p = 0.08958$ ), but gesture still is ( $\chi^2(10) = 26.585, p = < 0.001$ ). Speech results in an increase of character advantage of  $0.13 \pm 0.04$  (SE), whereas gesture results in an increase of  $0.18 \pm 0.03$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	-0.12560074	0.03584830	-3.504	0.0026
-speech;-gesture - -speech;+gesture	-0.18405292	0.03090626	-5.955	<.0001
+speech;-gesture - -speech;+gesture	-0.05845219	0.03741234	-1.562	0.4003
-speech;-gesture - +speech;+gesture	-0.26243383	0.04167290	-6.297	<.0001
+speech;-gesture - +speech;+gesture	-0.13683309	0.03090598	-4.427	0.0001
-speech;+gesture - +speech;+gesture	-0.07838090	0.03562682	-2.200	0.1232

Table 6.21: Character Advantage contrasts for Time Window 9

Table 6.21 reveals the same pattern as the window 8. However, in this time window, all conditions except -speech;-gesture are associated with positive mean scores for target advantage. It is likely that this is the result of participants returning the central fixation cross following selection of target item.

### 6.5.1 Summary of Results for Character Advantage

These results seem to strengthen the idea that participants continue to fixate on the character when unique information is presented in gesture but not in speech. Therefore this has a bearing on the finding that target advantage scores point towards the accuracy of hypothesis 2. While participants stopped fixating on the character earlier in those conditions where MANNER is presented in speech, the reason why they do not begin to fixate on the target in condition where MANNER is only presented in gesture is because they continue to fixate on the character. However, the data from time window three shows that this is not the case when no MANNER information is presented. Therefore, participants integrate crucial information presented in gesture very early on. The fact that participants are also more likely to reduce fixations to the character in the condition where MANNER information is presented in both speech and gesture, suggests that

gesture is not obligatorily processed but that comprehenders are selective.

The analysis of character advantage also helps explain the late effect of gesture shown the target advantage analysis. In this analysis from window 3 onwards conditions including gesture begin to account for increased character advantage and increased target advantage. This points to the fact that when MANNER is presented in speech (and when MANNER is not present) participants are likely to be looking at non-target and non-character items. The next section builds on these findings by exploring whether participants are more likely to be looking at the competitors, however, it is interesting to explore whether or not participants are fixating on the competitors which would suggest that participants are focussing on elements that are being disambiguated later or on.

## 6.6 Competitor Advantage

Competitor advantage is similar to character and target advantage, except looks to the distractor are subtracted from looks to the competitors. Therefore, competitor advantage depicts whether or not participants were more likely to look at the competitors than the distractors.

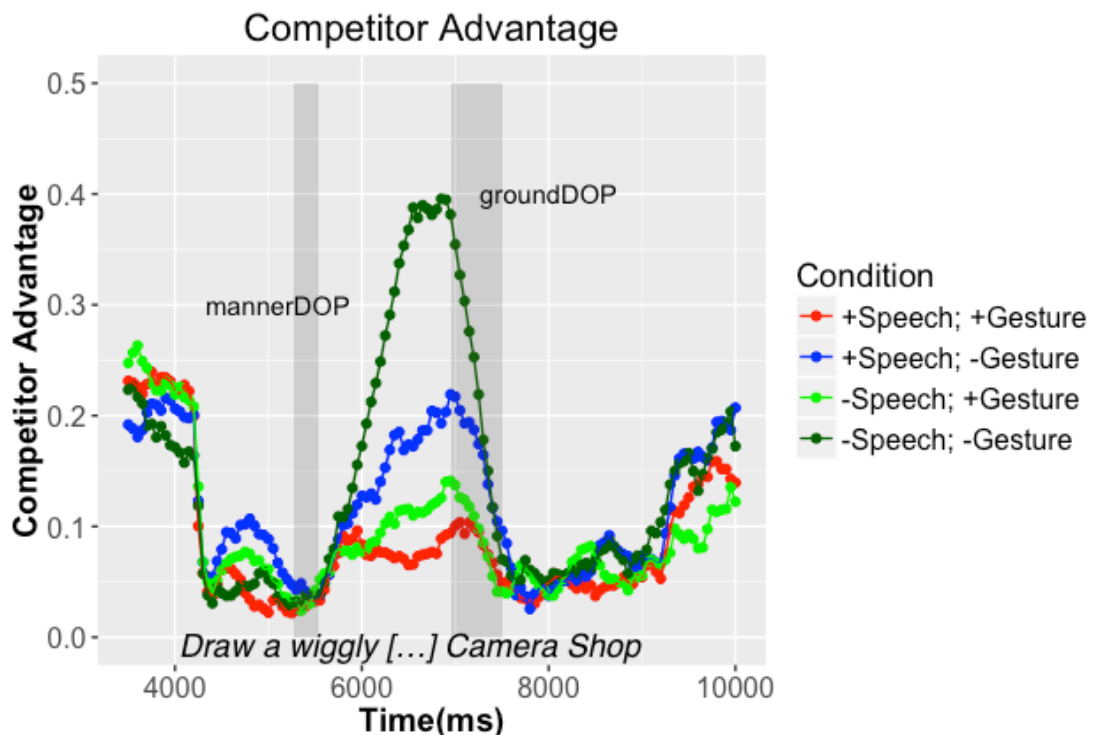


Figure 6.4: Competitor Advantage against Time

The following table shows the mean competitor advantage for each condition in each time window.

Condition	Mean Character Advantage (SE)
<b>Window 1: 5250-5550ms</b>	
<i>-speech; -gesture</i>	0.04 (0.01)
<i>+speech; -gesture</i>	0.04 (0.01)
<i>-speech; +gesture</i>	0.04 (0.001)
<i>+speech; +gesture</i>	0.3 (0.01)
<b>Window 2: 5550-5850ms</b>	
<i>-speech; -gesture</i>	0.09 (0.02)
<i>+speech; -gesture</i>	0.08 (0.02)
<i>-speech; +gesture</i>	0.07 (0.02)
<i>+speech; +gesture</i>	0.07 (0.02)
<b>Window 3: 5850-6150ms</b>	
<i>-speech; -gesture</i>	0.17 (0.02)
<i>+speech; -gesture</i>	0.13 (0.02)
<i>-speech; +gesture</i>	0.08 (0.02)
<i>+speech; +gesture</i>	0.10 (0.02)
<b>Window 4: 6150-6450ms</b>	
<i>-speech; -gesture</i>	0.29 (0.03)
<i>+speech; -gesture</i>	0.16 (0.03)
<i>-speech; +gesture</i>	0.11 (0.03)
<i>+speech; +gesture</i>	0.08 (0.02)
<b>Window 5: 6450-6750ms</b>	
<i>-speech; -gesture</i>	0.38 (0.03)
<i>+speech; -gesture</i>	0.19 (0.03)
<i>-speech; +gesture</i>	0.11 (0.02)
<i>+speech; +gesture</i>	0.07 (0.03)

**Window 6: 6750-7050ms**

<i>-speech; -gesture</i>	0.37 (0.03)
<i>+speech; -gesture</i>	0.21 (0.03)
<i>-speech; +gesture</i>	0.13 (0.03)
<i>+speech; +gesture</i>	0.10 (0.03)

**Window 7: 7050-7350ms**

<i>-speech; -gesture</i>	0.24 (0.02)
<i>+speech; -gesture</i>	0.18 (0.02)
<i>-speech; +gesture</i>	0.10 (0.02)
<i>+speech; +gesture</i>	0.09 (0.02)

**Window 8: 7350-7650ms**

<i>-speech; -gesture</i>	0.09 (0.02)
<i>+speech; -gesture</i>	0.10 (0.02)
<i>-speech; +gesture</i>	0.04 (0.02)
<i>+speech; +gesture</i>	0.05 (0.02)

**Window 9: 7650-7950ms**

<i>-speech; -gesture</i>	0.06 (0.02)
<i>+speech; -gesture</i>	0.05 (0.01)
<i>-speech; +gesture</i>	0.06 (0.01)
<i>+speech; +gesture</i>	0.05 (0.01)

The following table shows the results of the model comparisons for each time window.

---

	Competitor Advantage (SE)	$\chi^2$ (DF)	p-value
<b>Window 1: 5250-5550ms</b>			
<i>Speech</i>	0.008 (0.01)	0.0385(10)	0.8445

<i>Gesture</i>	0.004 (0.01)	0.3543(10)	0.5517
<i>Speech*Gesture</i>	-0.018 (0.02)	1.4773(12)	0.2242
<b>Window 2: 5550-5850ms</b>			
<i>Speech</i>	-0.007 (0.02)	0.1843(10)	0.6677
<i>Gesture</i>	-0.018 (0.02)	1.0739(10)	0.3001
<i>Speech*Gesture</i>	0.002 (0.02)	0.0079(12)	0.9293
<b>Window 3: 5850-6150ms</b>			
<i>Speech</i>	-0.04 (0.03)	0.1081(10)	0.7423
<i>Gesture</i>	-0.09 (0.03)	7.2383(10)	<b>0.007136</b>
<i>Speech*Gesture</i>	0.05 (0.03)	2.8831(12)	0.08952
<b>Window 4: 6150-6450ms</b>			
<i>Speech</i>	-0.13 (0.03)	7.4516(10)	<b>0.006338</b>
<i>Gesture</i>	-0.18 (0.04)	18.729(10)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	0.1 (0.04)	5.2502(12)	<b>&lt;0.02194</b>
<b>Window 5: 6450-6750ms</b>			
<i>Speech</i>	-0.19 (0.04)	10.594(10)	<b>0.001134</b>
<i>Gesture</i>	-0.27 (0.04)	31.332(10)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	0.15 (0.05)	10.404(12)	<b>0.001257</b>
<b>Window 6: 6750-7050ms</b>			
<i>Speech</i>	-0.16 (0.04)	7.39(10)	<b>0.006559</b>
<i>Gesture</i>	-0.24 (0.04)	25.975(10)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	0.13 (0.04)	8.4795(12)	<b>0.003592</b>
<b>Window 7: 7050-7350ms</b>			
<i>Speech</i>	-0.06 (0.02)	1.5661(10)	0.2108
<i>Gesture</i>	-0.14 (0.03)	22.297(10)	<b>&lt;0.001</b>
<i>Speech*Gesture</i>	0.05 (0.03)	2.4543(12)	0.1173
<b>Window 8: 7350-7650ms</b>			
<i>Speech</i>	0.004 (0.02)	0.5014(10)	0.4789
<i>Gesture</i>	-0.05 (0.02)	7.245(10)	<b>0.00711</b>
<i>Speech*Gesture</i>	0.0005 (0.02)	0.0005(12)	0.9818

**Window 9: 7650-7950ms**

<i>Speech</i>	-0.016 (0.01)	0.7273(10)	0.3937
<i>Gesture</i>	-0.0008 (0.02)	0.0004(10)	0.9834
<i>Speech* Gesture</i>	0.000005 (0.02)	0(12)	0.998

In this analysis, for reasons of space, rather than focussing on every time window, the focus will be on those time windows that have been critical in the other conditions.

**window 3: 5850-6150ms**

This analysis starts at window three because the for the first two windows participants were fixating on the character (as shown in the character advantage analysis). In this window there is a significant effect of gesture ( $\chi^2(15) = 23.184, p = < 0.001$ ) but not speech ( $\chi^2(15) = 23.184, p = < 0.001$ ). However, speech accounts for a reduction in competitor advantage of  $0.04 \pm 0.03$  (SE), whereas gesture results in a reduction of  $0.09 \pm 0.03$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	0.03893047	0.02693064	1.446	0.4708
-speech;-gesture - -speech;+gesture	0.08620974	0.02653215	3.249	0.0064
+speech;-gesture - -speech;+gesture	0.04727927	0.02934127	1.611	0.3721
-speech;-gesture - +speech;+gesture	0.07070059	0.03106843	2.276	0.1037
+speech;-gesture - +speech;+gesture	0.03177012	0.02652485	1.198	0.6283
-speech;+gesture - +speech;+gesture	-0.01550915	0.02668231	-0.581	0.9378

Table 6.24: Competitor Advantage contrasts for Time Window 3

Table 6.24 shows that the only significant difference is between the condition in which MANNER information is presented only in gesture and when MANNER information is not presented. This is due to the fact that participants are more likely to look at the competitors than the distractors in the no MANNER condition than when MANNER is conveyed through gesture. The average competitor advantage associated with this time

window shows that there is an continuous increase in competitor advantage going from -speech;+gesture, +speech;+gesture; +speech;-gesture to -speech;-gesture. This suggests that participant are more likely to look at competitors when the target is disambiguated with speech and not gesture.

**window 4: 6150-6450ms**

In this window there is a significant effect of speech ( $\chi^2(10) = 7.4516, p = 0.006338$ ) and gesture ( $\chi^2(10) = 18.729, p = < 0.001$ ). Both are associated with a reduction of competitor advantage, with speech reducing it by  $0.13 \pm 0.03$  (SE) and gesture reducing it by  $0.18 \pm 0.04$  (SE).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	0.13247731	0.03320331	3.990	0.0004
-speech;-gesture - -speech;+gesture	0.18064002	0.03445187	5.243	<.0001
+speech;-gesture - -speech;+gesture	0.04816271	0.03560435	1.353	0.5292
-speech;-gesture - +speech;+gesture	0.21314757	0.03762263	5.665	<.0001
+speech;-gesture - +speech;+gesture	0.08067026	0.03445442	2.341	0.0888
-speech;+gesture - +speech;+gesture	0.03250755	0.03284039	0.990	0.7552

Table 6.25: Competitor Advantage contrasts for Time Window 4

Table 6.25 shows that all conditions in which MANNER is presented result in significantly lower competitor advantage compared to the condition in which MANNER is not presented. This suggests that participants are using MANNER information in both speech and gesture to disambiguate the target from the distractors. Furthermore, the condition in which speech and gesture both present MANNER information results in significantly lower competitor advantage than when only speech conveys that information. However, gesture on its own is not significantly different from speech on its own. This time window suggests that speech and gesture together are a greater facilitator of looks to target than either gesture or speech on its own.



**window 7: 7050-7350**

In this window, speech does not significantly affect competitor advantage ( $\chi^2(10) = 1.5661, p = 0.2108$ ) whereas gesture does ( $\chi^2(15) = 22.297, p = < 0.001$ ). Speech reduces competitor advantage by  $0.06 \pm 0.02$ , whereas gesture reduces it by  $0.14 \pm 0.03$ .

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	0.06311608	0.02663517	2.370	0.0830
-speech;-gesture - -speech;+gesture	0.14043117	0.02749728	5.107	<.0001
+speech;-gesture - -speech;+gesture	0.07731509	0.02954478	2.617	0.0440
-speech;-gesture - +speech;+gesture	0.15221415	0.03130530	4.862	<.0001
+speech;-gesture - +speech;+gesture	0.08909807	0.02750650	3.239	0.0066
-speech;+gesture - +speech;+gesture	0.01178298	0.02638909	0.447	0.9703

Table 6.26: Competitor Advantage contrasts for Time Window 7

Table 6.26 shows that competitor advantage is not significantly different when only speech presents MANNER than when MANNER is not presented. Therefore, participants fixating on the competitors when only speech disambiguates the target with MANNER. Moreover, those conditions where gestures convey MANNER is significantly different from conditions in which it does not. Finally, gesture on its own is not significantly different from speech and gesture together. This window shows what that the speculation that when only speech disambiguates the target participants are less likely to fixate on the target and more likely to fixate on the competitors. This result is interpreted as suggesting that participants are surer of the target when it is conveyed through gesture than when it is conveyed through speech.

**Summary of Results for Competitor Advantage**

These results strengthen what has been suggested already. Participants who are presented MANNER information in speech are no less likely to be looking at the competitor items than when MANNER is not presented in speech or gesture. This suggests that the late difference between gesture and speech in terms of target advantage is the result of participants exploring the competitor items in the array. It is possible that gesture re-

sults in more stable representations of the target object based on the it being represented in the visual modality. This would follow from Wu et al.'s 2014, p. 49 suggestion that gesture promote “image-based representations” and the suggestion of Huettig and Altmann (2007) that visual representation are mapped onto memoried visual representations of elements on the display. Therefore, the visual nature of gesture might more strongly map onto the visual features in the display. However, before that hypothesis is explored it is necessary to rule out that it is not the case that participants in the +speech;-gesture condition are not searching the array because they have already selected the target object. If this were the case, then the lack of fixations on the target later in the task might be due to fact that participants who are given the information through speech are able to disambiguate the target object incredibly early.

## 6.7 Response Time

Response time represents the difference in time between the beginning of the trial and the point at which the participant clicks on the correct item in the array. For this analysis response time only includes correct responses and incorrect responses were removed. An analysis of correct vs incorrect responses could be the subject of future research. The following table shows response time in milliseconds by condition.

Figure 6.5 demonstrates that the difference between conditions is small and that participants are not selecting the target object early in any of the conditions. In fact, it looks as though participants, on average, wait until after the GROUNDPOD. In order to show these results in more detail, the following table presents the reaction times from slowest to fastest.

<b>Condition</b>	<b>Mean(SE)</b>
<i>-Speech; -Gesture</i>	7945 (53)
<i>+Speech;-Gesture</i>	7667 (78)
<i>-Speech; +Gesture</i>	7637 (61)
<i>+Speech;+Gesture</i>	7535 (91)

Once again a linear mixed effects analysis was conducted. The models used for this anal-

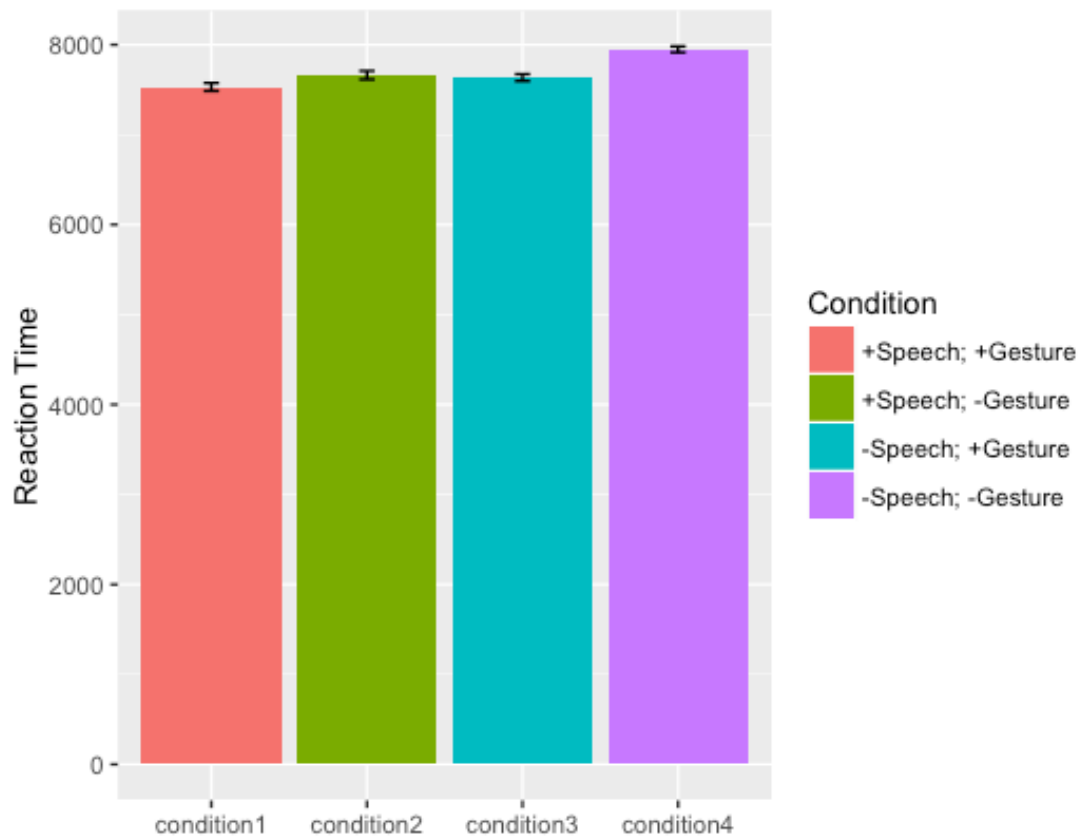


Figure 6.5: Reaction Time

ysis were the same as for the target advantage with the exception that response time is the outcome variable. Mixed effect analyses with speech and gesture as fixed effects found that gesture significantly predicted reaction time ( $\chi^2(10)=15.172$ ,  $p < 0.01$ ) accounting for a drop in reaction time of  $308.18 \pm 56.20$  (SE) but speech does not ( $\chi^2(10)=3.0423$ ,  $p=0.08112$ ), accounting for a drop in reaction time of  $278.58 \pm 67.08$  (SE). There was also a significant interaction between speech and gesture ( $\chi^2(12)=5.4427$ ,  $p=0.01965$ ).

Contrast	$\beta$	SE	z ratio	p value
-speech;-gesture - +speech;-gesture	278.57672	67.07580	4.153	0.0002
-speech;-gesture - -speech;+gesture	308.18304	56.66684	5.439	<.0001
+speech;-gesture - -speech;+gesture	29.60632	60.87417	0.486	0.9622
-speech;-gesture - +speech;+gesture	410.03303	78.47684	5.225	<.0001
+speech;-gesture - +speech;+gesture	131.45631	57.22277	2.297	0.0986
-speech;+gesture - +speech;+gesture	101.84999	66.78049	1.525	0.4223

Table 6.27: Reaction time contrasts

Table 6.27 shows that both speech and gesture significantly reduce reaction time compared to the no MANNER condition. However, there is no significant difference between the conditions in which MANNER information is presented in speech, gesture or both. However, the mean reaction times associated with each of the conditions shows that the order (in terms of lowest first) goes from +speech;+gesture, -speech;+gesture, +speech;-gesture, to -speech;-gesture. Therefore these results show that the late effect of gesture on fixations is not due to participants selecting the item earlier than in conditions where information is presented in gesture. Therefore, it is possible that gesture is associated with surety. Gesture results in participants being surer of which item in the display is the target.

In terms of the hypotheses outlined at the beginning of this chapter. Hypothesis 5, that gesture or speech will reduce reaction times, is correct. However, hypothesis 6 is incorrect since there is no advantage of speech. Hypothesis 7 also turned out to be correct (although not significantly) since speech and gesture, when presented together, are associated with the smallest reaction times. This final result also means that hypothesis 8, that presenting speech and gesture together will increase reaction times, is incorrect.

## 6.8 Discussion of results

These results present many interesting findings. First, there seems to be an early effect of MANNER in speech. This is demonstrated through the stronger effect of speech over gesture for predicting target advantage (although gesture is still significant) in the 5850-

6150ms time window. This finding was further demonstrated with character advantage in the preceding (5550-5850ms) time window because speech, without accompanying gesture, resulted in a significant reduction in fixations to the character. However, this was also the case in the condition where MANNER is not presented. This suggests that when speech presents MANNER information, participants immediately begin to reduce fixations to the character. In terms of character advantage in time window 3 (5850-6150) participants are less likely to be looking at the character when MANNER is not presented but continue to do so when it is. Therefore, participants continue fixating on the character when crucial information has been presented in gesture only. Utterances including gesture resulted in a significantly increased target advantage score within 275ms post MANNER offset. However, there was only an increase in terms of mean target advantage when gesture was also accompanied by speech conveying MANNER information. Although, when MANNER information is only presented in gesture, participants are more likely to fixate on the target than when MANNER information is not presented at all. This finding suggests two things. First, participants are able to extract MANNER information from gesture rapidly (within the time window that comprehenders process speech). And second, that there is an additive effect of speech and gesture. In the +speech;+gesture condition participants produced more fixations to the target in the 275 post offset window than in the +speech;-gesture condition.

These findings suggest that integration of speech and gesture occurs during the processing of lexical items, that comprehenders can extract information from gesture within the 275ms time window, and that speech and gesture together perform better than either modality alone.

There also seems to be a late effect of gesture. It was shown that in those conditions without gesture, participants begin to look around the array (at the competitors) beginning at about the 6450-6750ms time window. Importantly, it is during this time window when the DIRECTION and ORIENTATION element will be produced in speech. However, these elements could be extracted from the gesture earlier due to the simultaneous articulation of all semantic elements in gesture. This finding, therefore, has two interpretations. First, it seems that gesture provides participants with an early representation

of the target and therefore result in fewer fixations to competitors during the presentation of the DIRECTION and ORIENTATION elements in speech. And second, this suggests that the representation of the route shape that is derived from the gesture produces a more stable representation than the one derived from only hearing the MANNER component in speech. These results are interpreted in terms of Wu and Coulson (2014, p. 49) and Huettig and Altmann (2007). It seems that gestures promote image-based representations of the display items which are then mapped onto the representations of the display items that have been generated by looking at those items. It is possible that gesture maps more strongly onto the representation of the display items than those derived from speech. Therefore, when a participant disambiguates a display item based on MANNER information being presented in speech alone, they seem to doubt their developing representation as the utterance unfolds. With each new element of the spoken utterance participants fixate on the competitors that also depict those features. Crucially, this does not happen when both speech and gesture present MANNER information, which suggests that in this condition, participants get the best of both worlds. Speech provides the early disambiguation and gesture promotes image-based representations.

These findings, along with the map task results, will be compared to the general findings in the literature in the next chapter.

## **Part IV**

# **Discussion and Conclusions**





# Chapter 7

## Discussion

The purpose of this thesis is to explore gesture from the perspective of linguistic pragmatics. To do this, it is necessary to explain how gesture is incorporated into intentional-inferential model of communication (cf. Sperber and Wilson, 1995; Levinson, 1983; Levinson, 2000). Therefore, it is imperative to describe how gesture relates to the informative and communicative intentions of the utterance producer and the inferential comprehension procedure of the utterance comprehender. To this end, the preceding chapters have explored the effect of gesture at the different levels of the Clarkian action ladder (represented here as table 7.1).

<b>Level</b>	<b>Utterer A's actions</b>	<b>Addressee B's actions</b>
4	A is <i>proposing</i> joint project $w$ to B	B is <i>considering</i> A's proposal of $w$
3	A is <i>signaling</i> that $p$ for B	B is <i>recognizing</i> that $p$ from A
2	A is <i>presenting</i> signal $s$ to B	B is <i>identifying</i> signal $s$ from A
1	A is <i>executing</i> behavior $t$ for B	B is <i>attending</i> to behavior $t$ from A

Table 7.1: Action ladder involved in language use (Clark, 1996)

The Clarkian action ladder positions communicative acts within a series of simultaneously produced actions, all of which must be carried out for communication to be considered successful, or in Clarkian terms: “sufficient for current purposes”. Crucially, each level on the ladder is the product of the previous level and are therefore governed by the principles of upwards causality and completion and downward evidence (Clark, 1996, p. 152). In other words, as we go up the ladder, each action has to be produced in

order to satisfy the action at the next level. Moreover, each individual receives evidence that the previous action was successful because the next level has been reached. In this chapter, I begin by exploring the results of the map task and the visual world paradigm. These results will be interpreted in relation to dominant findings in the gesture literature. The second section will develop a theory of communication that emphasises the intentional nature of communicative gestures. The third section will defend this theory in relation to alternative theories in the gesture literature. This chapter will end with a discussion of potential future applications of the studies reported above.

## 7.1 Speech and gesture in joint actions

### 7.1.1 Signalling

Are speech and gesture composed to be a composite signals (Clark, 1996; Enfield, 2009b)? In other words, are both elements presented to be communicative? The results that address this question come from the production of utterances in map task, focussing on the information they convey and the effect of gesture on speech and speech on gesture.

#### Semantic Features

It was shown in chapter 4 that the semantic features distributed across speech and gesture were produced in a predictable way. The production of gesture seemed to be in a relationship with the presentation of *GROUND* information in speech. This relationship showed that as *GROUND* in speech increased, the presence of gesture became less likely. *GROUND* information is tied to the descriptions of map features, which were either given to the participants as part of the task (i.e., landmarks) or already established by the participants during the map task (e.g., a previously grounded section of the route). Therefore, *GROUND* relates to map features that are in within the common ground of the two participants. When turning to *DIRECTION* and *MANNER* information in speech (i.e., information that was directly tied to the route), it was shown in chapter 4 that while both *DIRECTION* and *MANNER* resulted in a significant decrease in the incidence of *GROUND* in speech, only

MANNER in speech was significantly affected by GROUND in speech. This suggests that descriptions of MANNER are less likely to occur with descriptions of GROUND. However, the incidence of DIRECTION in speech, when constrained by the presence of GROUND in speech, is not significantly reduced when compared to its average occurrence in the corpus. This suggests, therefore, that DIRECTION in speech occurs in the environment of GROUND in speech. This finding is important because it suggests that participants in map task describe landmarks and the direction of the route, but they are less likely to describe a landmark and the manner of the route.

When the particular semantic features of gesture were explored, it was shown that the presentation of both MANNER and DIRECTION information in speech had a positive effect on the presentation of both MANNER and DIRECTION information in gesture. This suggests participants are more likely to depict the MANNER and/OR DIRECTION of the route in gesture when they are talking about the MANNER and/OR DIRECTION of the route than when they are not. In other words, gesture reflects the semantic content of speech. It was also the case that MANNER in gesture increased the incidence of both MANNER and DIRECTION in speech. However, while DIRECTION in gesture increases the incidence DIRECTION in speech, it did not significantly increase MANNER in speech. Taken together, these findings suggest that speech describing DIRECTION is likely to occur with gesture depicting DIRECTION and MANNER but speech describing MANNER is no more likely to occur with gesture depicting MANNER but not DIRECTION. This suggests that participants describe MANNER without also depicting DIRECTION in gesture. In addition, when semantic features are explored within speech and gesture, MANNER and DIRECTION in speech are not significantly correlated, but MANNER and DIRECTION in gesture are. These results suggest that because of the coupling of MANNER and DIRECTION in gesture, when individuals describe DIRECTION in speech the MANNER comes ‘for free’ in gesture.

These results suggest that describing elements in common ground is more likely to occur in environments where participants do not gesture than environments in which they do. In other words, features that are grounded or perceptually co-present are less likely to be accompanied by gesture. This finding is in line with the literature suggesting that the incidence of gesture decreases (or is reduced in size and precision) with

grounding (Holler and Stevens, 2007; Gerwing and Bavelas, 2004; de Ruiter, Bangerter, and Dings, 2012b; Jacobs and Garnham, 2007). Furthermore, descriptions of `GROUND` information are not significantly less likely to occur with `DIRECTION` information, suggesting that perhaps participants describe the direction the route is travelling in relative to features that have been grounded. This is not the case for manner described in speech, which suggests that the `MANNER` of the path is described separately from the shared features. Therefore, it is possible that participants are adopting different communicative strategies. These can be listed as follows:

- establish landmarks through speech
- describe the direction of the route relative to landmarks through speech
- describe the route's manner or direction whilst producing gesture

Importantly, the gesture tends to conflate `MANNER` and `DIRECTION` information. Therefore, it is possible for gesture to provide additional information not necessarily described in the speech. Furthermore, information relating to `MANNER` that was depicted in gesture almost always was the product of a tracing gesture. However, `DIRECTION` information could be depicted with a tracing gesture or a directional gesture (e.g., moving a flat hand in the direction the route is travelling). These findings suggest that gestures depicting the route with a tracing gesture are more likely to occur with speech describing the manner whereas those only depicting the direction of the route are not. This suggests that it is the conceptualisation of the route, as a route, rather than as abstract direction, that triggers the production of tracing gestures. It is the strategy adopted by the utterance producer, as well as the semantic content of the speech, that is crucial for whether or not an utterance contains gesture.

In line with previous literature, people gesture more when the information covered through gesture is of high importance (Beattie and Shovelton, 2006). Moreover, while gestures are semantically linked to the meaning of the speech they accompany, they can depict information not in speech (Cohen, Beattie, and Shovelton, 2011). People also seem to gesture less when describing nameable, shared features than when describing unshared route features. This could either be because those nameable features are not

spatial (in line with the findings of Wesp et al. (2001)) or because shared features are in the participants common ground. What's more, these result suggest that participants produce gesture in a predictable way, it is not random.

### **Perspective**

Semantic units in the map task are divided into five categories based on the perspective adopted in gesture by the utterance producer: 1. no gesture, 2. manual behaviours that are produced for the utterance producer, which their addressee can not see (e.g., tracing the route directly on the map), 3. first person gesture, which are produced from the producer's perspective, 4. second person gestures, which are produced from the addressee's perspective, and 5. shared perspective gestures, which involve the utterance producer turning so that they share gesture space with their addressee. The results associated with perspective are crucial, because the use of different perspectives in the map task can be thought of as acting like a natural version of the mutual visibility studies (Bavelas and Healing, 2013). However, rather than the experimenter manipulating mutual visibility, participants decide which manual movements to show their interlocutor. The first finding demonstrated that there was a relationship between whether GROUND was described in speech and the perspective taken by the producer. As discussed above, the presentation of GROUND in speech was negatively correlated with the production of gesture. It was demonstrated that gestures produced from a first person perspective affect the amount of GROUND in speech. However, manual movements produced beyond the view of the addressee do not significantly reduce the amount GROUND in speech. In other words, participants seem to be sensitive to whether their addressee can see what they are doing with their hand, even if they do not explicitly adopt their addressee's perspective.

Moreover, turning to those perspectives where gestures were visible to the addressee, it was shown that there was no a difference in terms of DIRECTION and MANNER in gesture. In other words, gesture seems to convey the same information in the three categories. Another finding was that speech-framed gestures (i.e., gestures that are referred to by the utterance producer) occurred more with gestures that adopted the addressee's

perspective. This suggests that participants are more aware of their gesturing when it is produced from a shared perspective.

### **Motion**

The presence of motion (e.g., utterances including the verb “goes”) in speech was positively correlated with gesture and negatively correlated with GROUND in speech. Therefore, descriptions representing the route as if it is moving are more likely to be accompanied by gesture. This finding can be thought of from the perspective of studies showing that the production of gesture is tied to competing representations (Kita and Davies, 2009). Since the route on the map is static, but the route (as it is to be drawn by the follower) is in motion, it is possible that the giver has competing representations of the route. It could be this competition that increases the use of gesture.

### **Complexity**

It was shown (section 4.2.8) that gesture is negatively correlated with semantic complexity in speech. In other words, the more semantic features in a single semantic unit, the less likely it was that gesture would be produced with that unit. However, since the only elements that can appear multiple times are GROUND, FIGURE, and POSITION, then the more complex semantic units are usually more complex because they include reference to multiple items on the map. This goes in line with the view being developed that speakers use landmarks as anchors for the route and, when they do, they are less likely to produce gesture. In other words, while a gesture can provide a dynamic representation of a route, anchored to gesture space, speech without gesture must decompose the route on the map into a series of coordinates that can be joined up.

### **“Around”**

The production of gesture in the environment of the word “around” was explored. The word “around” is underspecific, since it does not specify the direction of movement. In the context of “around” it was shown that the presence of gesture does not affect the presence of GROUND in speech, nor did gesture significantly affect the presence of DIRECTION in speech. In other words, unlike what is found generally, gesture is no less

likely to occur with GROUND in speech and no more likely to occur with DIRECTION in speech. Furthermore, DIRECTION in speech is not significantly affected by the presence of “around” suggesting that it is not the case that participants are specifying the direction of the route with speech. However, the presence of “around” does significantly increase the presence of DIRECTION in gesture. In other words, it seems as though gesture is performing a significant role in the specification of “around” This suggests that gesture is being used to reduce the potential misunderstanding that a comprehender might have when comprehending underspecific lexical items.

### 7.1.2 Recognising

For Clark, recognising is the process of comprehending the meaning of an utterance. The eye tracking study was designed to access the various effects speech and gesture have on the real time comprehension of utterances.

The results of the fixation analysis (Chapter 6) revealed an early advantage for disambiguating the target when information pertaining to the MANNER of the route is presented through speech. Importantly, this advantage was only seen when MANNER information was not conveyed through gesture. However, it was shown that when MANNER information was only presented in gesture, participants were actually more likely to fixate on the character in the video stimulus than anywhere else. Therefore, what this finding actually shows is that people are able to distinguish meaningful gestures from those not conveying meaningful contributions. This finding is in line with studies that have demonstrated an earlier priming affect of speech over gesture (cf. So et al., 2013, p. 779, non-significant result) and that speech is processed faster than gesture (Kelly, Özyürek, and Maris, 2010). It is possible therefore, that there is an advantage of processing symbolic language over iconic gesture. However, it is also possible that participants look at the video for longer because they are expecting additional task critical information. This could be tested in future studies by exploring differences in priming for speech, iconic gestures, emblematic gestures, and sign language. If there is an advantage of symbols then both emblematic gestures and sign language should result in earlier priming effects than iconic gesture (analogous to the priming effect of speech). If it is

tied to modality, then speech should still be processed faster than other communicative behaviours. Future studies need to distinguish between semiotic ground and modality.

It was also demonstrated that as the utterance unfolded, the conditions in which MANNER information was presented in gesture resulted in a greater proportion of fixations to the target compared to those in which gesture did not convey MANNER. There are two reasons why this might be the case.

The first is that the iconic nature of gestures (i.e., visual representations of visual referents) creates a stronger link than the symbolic nature of spoken referents. If this is the case then the continual rise in looks to target might be due to the fact that the comprehender is surer which referent is being described. This view is in line with the findings of Huettig and Altmann (2007) who found that spatial representations that come from utterances are mapped onto memorised representations of items in the visual world. The suggestions of Huettig and Altmann (2007) can be compared to the suggestion that gestures “promote image-based simulations of the meaning of an utterance” (Wu and Coulson, 2014, p. 49). Therefore, perhaps the iconic nature of the spatial representations in gesture map onto the memorised representations of map elements, strengthening the activation of linguistic meaning. In other words, gestures are more strongly mapped onto the memorised representations of the images in the visual world.

The alternative view is that speech activates representations quickly, but the linear nature of spoken utterances means that comprehenders have to retain the emerging representation associated with sequential lexical items in working memory. As each lexical item is comprehended, the comprehender updates their representation. Therefore, while the correct referent is disambiguated early on, they are more likely to challenge that representation as non-disambiguating lexical items emerge in the utterance. For example, if a participant heard the utterance “Draw a wiggly line going right under the parked van”, they would be able to disambiguate the target on the word “wiggly”. However, the words “right” and “under” also disambiguate the competitor items. It seems as though participants look at the competitors when they hear these later disambiguated words. This is not the case when gesture depicts a wiggly MANNER. This view was evidenced by the fact that when MANNER information was only presented in speech, participants



were as likely to fixate on the competitors (which shared direction and orientation features with the target) as they were in the condition where manner information had not been provided. However, this was not the case in the conditions where gesture depicted MANNER information.

A further explanation of this late effect, might be tied to the fact that participants were very quick to stop looking at the character when the gesture did not depict MANNER information. It is possible that participants assumed that all gestured information was not meaningful. In other words, they could not rely on gestured information so they ignored it. If this was the case, it is possible that the participant misses the DIRECTION and ORIENTATION features that occur earlier in gesture than in speech, because the semantic features depicted in gesture are produced simultaneously. However, it still suggests that participants are not necessarily strengthening their representation with each successive lexical item, but that non-disambiguating lexical items that occur later in the utterance cast doubt on already disambiguated referents. Therefore, it suggests something crucial about the condition in which manner is represented in both speech and gesture. Participants do not only pay attention to one modality, they pay attention to each modality which mutually strengthen each other. The potential doubt that comes as a result of having to retain information in working memory is alleviated by the presence of gesture.

Finally, in terms of reaction time, correct selection was faster when information was presented through gesture (section 6.7). However, this finding was tied to the added advantage of the condition where MANNER information was presented through speech and gesture. It was demonstrated that the difference between utterances that presented MANNER information in speech on its own and gesture on its own was not significant. Therefore participants selected the correct item as early when they received crucial information in speech or gesture.

### 7.1.3 Proposing

Proposing is determined, not by the meaning of an utterance, but by the action an individual is proposing in using that utterance. Very often, such acts are called ‘speech

acts' (see Kissine, 2013, for a recent treatment of speech acts). For example, in the map task the same words (i.e., “you go round the pyramid”) could be an instruction, a check for clarification, or a clarification of a previous utterance. Proposing was explored by investigating the move-types used in the map task. However, in this thesis much less time has been spent exploring proposing (and considering). These are clear areas of expansion for future work. It was shown that moves requesting information (checks) are associated with reduced gesture vs. moves that convey information (clarifications and instructions). Therefore, it seems likely that gesture is tied to the presentation of information for an addressee rather than the request for information from addressee. This seems like a rather straightforward finding. It would be impossible to gesture what something looks like when you do not know what it looks like. However, this view cannot explain why it was also demonstrated that instructions were associated with more gestures than explanations. Explanations are unsolicited descriptions of map features that do not constitute instructions. In other words, explanations are moves that relate to features on the map that are shared (e.g., landmarks) rather than unshared (e.g., the route). This finding therefore points to the fact that moves that convey task crucial information are more likely to occur with gesture than those that attempt to coordinate location. This goes in line with findings showing that important information is more likely to be communicated by gesture (Beattie and Shovelton, 2006).

#### 7.1.4 Considering

Considering is the final element on the Clarkian action ladder and it relates to the uptake of a proposal. For example, if the proposal took the form of an instruction then the response might be an acknowledgement that the instruction has been understood this would close the joint project initiated by the proposal. However, it might be the case that a proposal is not understood (or the addressee is unable to carry it out). In this case they may ask for clarification, which would not close the project but begin an embedded project. Considering was explored in the map task by investigating the presence of gesture in successively deeper joint projects. The depth of a project was determined by whether or not the current project related an ongoing project or began in a new

project following the closing of the previous one. Projects are not closed until both participants are satisfied that the project is complete. Projects are continued when either participant continues the project, either by elaborating on previous given information or by asking for clarification. It was shown that as participants went deeper into projects, they gestured less.

There are several potential answers as to why this might be the case. The first is that when participants elaborated on something they have already said, either because clarification was requested or because they simply decided to do so, they gesture less because they are consciously attempting to be clearer. In common parlance, there is a suggestion that gesture is not for substance (McNeill, 2015, p. 4), and perhaps less clear than speech. Therefore, it is possible that conscious attempts at clarity are coupled with a reduction of gesture. However, this runs counter to the literature suggesting that speech with gesture is generally clearer than speech occurring on its own (Hostetter, 2011). A second possibility is that the clarifications and elaborations are about referents that have already been established (although not satisfactorily). If this were the case, the reduction of gesture could be tied to the fact that the things being described are (semi-) grounded. The final possibility is that there are participants who are more likely to require embedded projects to explain a section of the map and that those individuals are less likely to gesture. In other words, they are less clear on the first attempt and need to provide additional information. If this is the case then it suggests that gesture is associated with successfully getting your point across at a higher project level. This final position is consistent with (Bavelas et al., 2011), who showed that information conveyed through speech with (non-redundant) gesture are more efficiently grounded than speech on its own. Regardless of which view is taken, these results do suggest that gesture is having an effect at the level of the ongoing project. This second perspective seems more compelling since it is based on how people actually behave. However, future work could further explore the data by investigating, not just level, but the particular moves at each subsequent level.

### 7.1.5 Summary of main findings

Bringing these findings together, it seems that gesture does have an effect on every level of Clark's action ladder. Gestures are distributed predictably in terms of signalling (i.e., meaning), they have an effect on the recognition of signals (i.e., comprehension/understanding), they are determined by the proposal of the utterance producer, and they affect the how a proposal is considered in terms of the level at which a project reaches before it is closed. Therefore, this suggests that theories of communication necessarily need to be capable of incorporating gesture as a fundamental aspect of communication. Following Clark's model then, we are in a position to say what gesture is doing. Gesture plays a role in the upward activation of levels in the action ladder, and it is taken as evidence that a previous level has been achieved. However, what Clark's view does not explain is why a particular communicative behaviour was produced rather than another one. Why, at the level of signalling, did an utterance producer choose that particular composite signal over another one. For example, why does one person produce speech and gesture, when another only produces speech? He offers the view that signal choice is based on purpose, availability, and effort (Clark, 1996, pp. 186f.), but these terms are incredibly difficult to explore. The key objective of this thesis is to explore gesture from the perspective of pragmatics. A pragmatic perspective must explain the relationship between gesture and the intentional-inferential communication system (cf. Sperber and Wilson, 1995; Levinson, 1983; Levinson, 2000). Therefore, it is crucial to explain why speech and gesture are composed in the way they are and not just that they have an effect. There are two distinct perspectives in the pragmatic literature concerning the nature of gesture as part of the communicative system. The next section argues for a model that properly incorporates the findings from the studies presented above and gesture research more generally.

## 7.2 Gesture and Pragmatics: the predictability of gesture

### 7.2.1 An alternative model of gesture production

The two dominant views of gesture in the pragmatic literature are as follows. Wharton's (Wharton, 2003; Wharton, 2009) builds on a relevance theoretic view of communication, whereby all behaviours have the potential to be part of the relevance theoretic comprehension procedure (Sperber and Wilson, 1995). Linguistic signals are communicative because human cognition is geared towards processing them. Gestures, on the other hand, are communicative because they are taken as shown natural behaviours. In other words, gestures have meaning because they have meaning naturally. For example, a cough might naturally mean that someone is unwell, and therefore a producer can bring an addressee's attention to the fact they are unwell by emphasising the cough. If gestures work in this way then it is necessary to explain what their natural meaning is. Wharton (2009) suggests that they are behaviours geared towards helping in speaker understand what they are saying, and thus communicative because they have this purpose. This view leads to two assumptions that have an effect on a model of gesture production. First, gestures are not generated by the producer's informative and communicative intentions, and second, gestures can communicate the same content regardless of whether they are produced to communicate. Wharton's view, however, does not fully explain why participants pay attention to the type of iconic gestures that are the focus of this thesis, but instead treats all gestures in the same way. Therefore, while Wharton's view includes gesture in communication, it performs a secondary role to speech. At the very least the communication of gesture and the communication of speech are treated as distinct.

The alternative perspective, which is most elaborately explained by Enfield (2009b) and Enfield (2013), builds on Clark's view, and suggests that speech and gesture jointly constitute utterances. Speaking in terms of communicative intentions, this means gesture and speech are both part of the speaker's communicative and informative intentions. This is built on the idea that meaning in language is derived by the same princi-

ples as meaning elsewhere, because it is a public element of a cognitive process (Enfield, 2013). Gesture is related to referents in ways distinct from speech, but together they are related to referents in a composite fashion, and the relationship between the different components of a sign and the referent is greater than the sum of the two parts. Enfield suggests a series of heuristics that guide comprehension, so that both speech and gesture are comprehended together. However, Enfield's perspective only explicitly describes comprehension, what he calls *sign filtration*, and therefore does not explicitly present a theory of how speech and gesture are produced together. We are therefore left with Clark's (1996) view that the choice of composite is based on purpose, availability, and effort. Therefore, this view does state that speech and gesture are meaningful in the same way. It is worth reiterating that purposes are non-intentional (cf. Kockelman, 2005). A screwdriver can have a purpose but it cannot have an intention. This is not a trick of English where the purpose of the screwdriver is the purpose of the manufacturer or owner, because it's also possible to say that a rhino's horn has a purpose even though it was not created by anyone. Intentions are special because they connect mental states to the world (Searle, 1983; Dennett, 1993).

Comparing these views to those in the gesture literature, Wharton's view is most consistent with Krauss's lexical access view (Krauss, Chen, and Gottesmann, 2000), since gestures can be used to communicate but communication is not their natural function. Enfield's (2013) composite utterance view has been compared to the trade-off hypothesis by De Ruiter (2007). The trade-off hypothesis is built on Clark's perspective, but Krauss, Chen, and Gottesmann (2000) have criticised it for not being true to Clark's collaborative vision of language use. However, (Enfield, 2009b; Enfield, 2013) does not explicitly describe the cognitive processes that underlie speech and gesture production, but the suggestion seems to be that, although semiotically distinct, speech and gesture are a product of the same inferential communicative process. It is important that both Wharton and Enfield suggest that gesture will play a role in comprehension. However, only Wharton explicitly provides a model of gesture production. In summary, there seem to be two types of views on gesture production within pragmatics: those that consider gesture to be communicatively and informatively intended and those that consider

gesture to be a product of natural behaviours that are incorporated into communication by being deliberately shown. The studies presented in this thesis suggest that gesture is produced as a fundamental part of utterances, which is produced alongside speech to communicate. In other words, both speech and gesture are generated as part of a producer's communicative and informative intentions. This suggests that Wharton's view (and Krauss's), since it treats gesture as a distinct behaviour cannot entirely capture gesture's role in communication. Gesture is not simply derived from another process, gesture is often generated with the purpose to communicate. Therefore the only view that efficiently captures gesture from a pragmatic perspective is Enfield's composite utterance view. Enfield's view explains the complex process of comprehending speech and gesture composites. However, the details regarding composite utterance production are left unexplained. In the rest of this section a model of production will be outlined before being compared to current models of gesture production in the literature.

The perspective taken in this thesis is that gesture is produced communicatively and is part of the informative and communicative intention of the producer. In other words, the behaviour a communicator produces is not simply a coded message that can be decoded by the comprehender, but it is a behaviour that is causally tied to the producer's intention in producing it, such a view has been referred to as an ostensive-inferential perspective on communication. Therefore, the goal of the producer is to create a situation in which a comprehender can infer the intention underlying a particular behaviour (and know that behaviour was produced for that reason). One of the key shortcomings of the views of production discussed in the gesture literature is that they treat the production of an utterance as the goal of the communicative process (cf. Kita and Özyürek, 2003; Krauss, Chen, and Gottesmann, 2000; de Ruiter, 2000). Therefore, the end point of the productive process is the signal itself. Theories of production that posit signal production as the end product of the process miss a key point highlighted by pragmatic perspectives on language use. They tend to focus on the information as it is distributed across the modalities, giving rise to suggestions that in certain situations gesture can be used redundantly. However, the measure of an utterance is not only how much information it conveys (for the analyst), but also how it is dealt with by the comprehender.

Before elaborating further on the effect of such a view, it is worth spending some more time on the exact nature of the intentions in communication.

It was suggested in chapter 2 that while pragmatic scholars focus on how the intentions found in communicative acts differ from the intentions found in non-communicative acts, less time is spent actually defining what an intention is for the speaker. This lies in the fact that pragmatic models are often based on inferences from the effect of a behaviour to the intention underlying it. Reproducing the example from chapter 2:

(7.1) **Bob** How was your day?

**Anne** God, I need a drink

+> Not Good

Using example 7.1, the intentional-inferential process would likely be described as follows. The effect of Anne’s utterance is that it makes Bob think that Anne has not had a good day. Since Anne is able to meta-represent Bob’s mental state (via mindreading), then she intended Bob to think that she had not had a good day, and she did this by saying “God, I need a drink”. This seems like a fairly robust explanation of what is happening here, however it has a flaw. It is initially based on Bob’s view of the situation and thus begins with effect and develops a theory of behaviour. We may think of such views as product driven perspectives on inferential communication. The alternative, following Enfield (2009a), may be referred to as imperative driven perspectives on inferential communication. In what follows, I will try and build a case for why an imperative driven perspective is crucial for understanding why people choose to gesture.

For Searle (1983), an intention, like other forms of intentionality, has three key features. The first two of these are direction of fit and direction of causation. To describe “direction” Searle uses the terms “world to mind” and “mind to world”. In terms of an intention, it can be said that the direction of fit is world to mind and the direction of causation is mind to world. For example, if I intend to close a window, I change the world (in which the window is open) so it fits my mental representation (in which it is closed), as such the world is changed to match my mind. This is the direction of fit. However, the causal relationship goes from the mind to the world: it is because I have an



intention to close the window (in my mind) that I perform an action to close the window (in the world). This raises the question of what satisfies intentions. In the case of closing the window, the intention is satisfied once I have closed it. This feature, Searle calls the intention's conditions of satisfaction. However, intentions are causally self-referential. For an intention to be satisfied, the intention must be part of the process of achieving its conditions of satisfaction. If I intend to close the window, but someone else has done it, then my intention is not causally responsible. My intention is left unsatisfied. This distinguishes intentions from desires. For example, if I desire that the window be closed, the direction of fit is the same but desires are not conditionally self-referential. If someone else closes the window then my desire is still satisfied.

This view, so far, is rather simplistic. If I intend to close a window, it is not enough to simply say that I do it. In order to satisfy my intention I might need to get up, walk over to the window and pull the window shut. It is possible to go on dividing up intentions indefinitely, in what has been referred to as the “accordion effect” (Searle, 2010, p. 37). In this sense the intentions involved in an act can be compressed or expanded. This can be described using Clark's action ladder, for example in the map task where an individual says “draw a line going around the right of the pyramid”, they are instructing their addressee. However, if we expanded the ‘accordion’, then we can shift our focus to those actions lower down the ladder (e.g., signalling). The “accordion” effect can therefore be thought of the general principle governing what Clark is addressing. Returning to the action of opening the window, crucially, getting up and the other acts performed as a result of my intention to close the window are also the product of intentions. Furthermore, opening the window might be an action performed in order to satisfy another intention (e.g., lowering the temperature of the room or stopping the smoke alarm going off). To distinguish between the two types of intention, Searle uses the terms “prior intentions” and “intentions-in-action”. So, if I form a prior intention to close the window, then in order to satisfy that intention I must form a series of intentions-in-action, each of which has its own conditions of satisfaction (but is geared towards to satisfaction of the prior intention). If I want to stretch my legs and I stand up, someone might ask me “Why did you stand up?”, to which I could reply “To stretch my legs”. However, if I have formed a

prior intention to close the window and standing up is an intention-in-action, then my answer to the previous question would be “To close the window”. Most people would not think that I believe standing up automatically closes the window, but that closing the window requires one to get up and walk over to it. The general point highlighted by this perspective is that the difference between intentions-in-action and prior intentions is often one of perspective. Kissine (2013, p. 43) explains this phenomena when he states:

To use a somewhat abusive formulation, intentions ‘store information’ about future events. Not objectively, of course, since we may intend to do things that never get realised, but, from the subjective point of view, the content of an intention is a future event to the achievement of which intentions-in-actions are geared.

The structure of prior intentions and intentions-in-action is represented by Searle (2010) as follows:

(7.2) Prior Intention  $\rightarrow$  (Intention-in-action  $\rightarrow$  Behaviour)

In this formula the behaviour which satisfies the prior intention is caused by an intention-in-action, which is itself caused by the prior intention.

While the accordion effect means that we are able to shift our perspective and focus on different prior intentions and intentions-in-action, the focus here is specifically on the informative and communicative intentions. Informative and communicative intentions can be treated as a prior intention and an intention-in-action respectively. So, for example, if I wanted to inform someone that it is raining (my informative intention), then I could do that by saying “it is raining”, by opening the curtains, or by pointing out the window. All these different communicative behaviours are the result of a communicative intention. It can be argued that the communicative intention is an intention-in-action for the informative intention, which is a prior intention. Therefore, the numerous behaviours performed in order to satisfy the informative intentions will each be produced in response to different intentions-in-action. The structure of multiple intentions-in-action can be represented as follows:

$$(7.3) \text{ Prior Intention} \rightarrow \left( \begin{array}{l} (\text{Intention-in-action 1} \rightarrow \text{Behaviour 1}) \\ (\text{Intention-in-action 2} \rightarrow \text{Behaviour 2}) \end{array} \right)$$

Therefore, turning to just the spoken behaviour and pointing behaviour from the example above, it is possible to formulate the process as follows:

$$(7.4) \text{ Inform John that it is raining} \rightarrow$$

$$\left( \begin{array}{l} (\text{Say: "John, it's raining"} \rightarrow \text{"John, it's raining"}) \\ (\text{Direct John's attention to the window with a point} \rightarrow \text{Point at the window}) \end{array} \right)$$

At this point it is important to clarify the distinction between simultaneous and sequential acts. For example, if I want to close the window, first I have to get up and then walk over to it. I cannot do this the other way round. In Clarkian terms, these are two acts performed sequential that are both part of the larger activity of closing the window. However, it is also the case that in order to get up, I need to use both my left leg and my right leg, two acts that are simultaneously performed as part of the act of getting up. The production of speech and gesture together, is a simultaneous and not sequential act. To clarify, if I intend to inform you that it is raining, then I must have an intention-in-action geared towards doing that. One of the key arguments of this thesis is that the different behaviours one can employ in achieving the conditions of satisfaction of an informative intention are selected on the basis how likely it is that those different behaviours will satisfy the informative intention. Elaborating on this point, it is possible that the intention to communicate with gesture and the intention to communicate with speech are both (simultaneously articulated) intentions-in-action geared towards a prior intention. If this is the case then the condition of satisfaction for those intention-in-actions are that the two components successfully communicate what the speaker intends to communicate. However, there is also a third assumption that must be made. Consistent with Searle's view, Kockelman (2012, p. 12) states:

an intention is not just causal of a state of affairs, but also in need of a reason, its satisfaction conditions may include the belief (and perhaps pro-attitude)

which justifies it. In other words, an intention may be the conclusion of a practical inference: 1) if I open the door, then I can enter the room; 2) I want to enter the room; 3) so I shall open the door. Such an inference has premises: a relatively foregrounded conditional (a belief involving an if-then sequence); a relatively backgrounded pro-attitude (qua desire, status, or value). And such an inference has a conclusion: the intention itself (I shall open the door). If asked to provide a reason for one's behaviour, one may articulate such a sequence: both a belief (if-then) and a pro-attitude (a desire, status, a value).

Kockelman's view presents the imperative driven perspective on intentional acts. In other words, it is because the person wanted to enter the room, that they opened the door and not that they opened the door because they wanted to enter the room. Using Kockelman's formula with communicative acts, it is possible to say 1) if I produce behaviour X, then John will know/do Y; 2) I want John to know/do Y; 3) so I shall produce behaviour X. If we assume that behaviour X is a complex behaviour that may include speech and gesture, then the central question for a theory of gesture production, is how do producers arrive at beliefs/assumptions such as "if I produce behaviour X, then John will know/do Y"? In order to provide a theory that can answer such a question, it necessary to turn to a decision theoretic view of communication. Here, I do not want to fully embrace a game theoretic model, but only to borrow useful terminology (cf. Benz, Jäger, and van Rooij, 2006). It is possible to define the *utility* of an intentional action as the extent to which that action satisfies the conditions of the intention minus the effort it takes to perform it. For example, if I form a prior intention to close the window, the utility of my intentions-in-action can be defined in terms of whether the window actually gets closed and how much effort it takes to perform that action. For example, if I walk over to the window and close it, then that satisfies my prior intention, but so does cartwheeling over to the window and closing it. The main difference is that the effort expended in cartwheeling (compared to walking) reduces the utility of cartwheeling. While this is a fairly simple case, it introduces a problem. Kockelman suggested that a prerequisite for performing an intentional action is the justifying belief that that

action will satisfy the intention. The question we asked at the start of this paragraph can now be changed to “How does an individual know the utility of an action before they perform it?”. The answer is that they do not know the utility of an action, they can only estimate the utility. Such estimations Benz, Jäger, and van Rooij (2006) call *expected utility*. Expected utility can be thought of as actual utility plus error. It seems fairly obvious that cartwheeling will expend more effort than walking, so someone faced with a decision of deciding whether to walk over to window or cartwheel over to the window, is able to easily estimate the utility of walking compared to cartwheeling. In other words, it is likely that there would be minimal error in working out the expected utility of cartwheeling and walking.

How does this work for communicative behaviours? Imagine I want to inform John that it is raining. My prior intention has the conditions of satisfaction that John knows it is raining (or, the very least, John knows I believe it is raining). My intention-in-action results in the behaviour I produce that allows John to infer the fact that I want to inform him that it is raining. To simplify things, imagine I have three choices of behaviour: 1) say, “John, it is raining”; 2) point out of the window; or 3) say “John, it is raining” and point out of the window. First off, what is the utility of each action? This question is far less easy to answer than the previous question regarding utility. Therefore, we can assume that, for an utterance producer, there is more error associated with this estimation than the previous one. However, on the assumption that 1 and 2 both satisfy the condition of the prior intention to inform John that it is raining 3 has less utility (based on the effort required to produce two behaviours rather than one). However, is this assumption valid? How would an individual know that 1 and 2 will both satisfy the conditions of their prior intention? They would have to estimate the utility of each action and then compare them to one another. As actions become more complex and the number of distinct options to satisfy a prior intention are increased, the decision between different intentions-in-actions become increasingly untenable. In other words, increasing the complexity of an action will increase the number of options open to an individual and the amount of error associated with any choice. Working out which behaviour is correct in light of the increasing difficulty of a decision to be made will take effort. If this is

the case then the effort required to make such estimations would have to be part of the model. Furthermore, it is important to point out that the decisions in these examples are incredibly simplistic compared to the decisions people make in genuine communicative situations, which have a much greater number of possible behaviours. Therefore, the answer proposed here is that for most cases, it seems unlikely that individuals work out the expected utility of the numerous potentially communicative behaviours open to them. If this is the case then how does a producer decide which communicative behaviours to produce? The suggestion of this thesis is that because estimations of utility are difficult, individuals do not calculate utility, but develop heuristics that shortcut the decision making process (Gigerenzer, Todd, and ABC Research Group, 1999). One heuristic, which we might call the *predictability* heuristic, states: all else being equal, perform the action for which expected utility is most likely to satisfy the conditions of the prior intention. One might question what is the difference between expected utility and predictability. Whereas expected utility is an estimation of an action satisfying the conditions of a prior intention minus the effort it takes to do so, predictability also takes into account the effort required to make that estimation. So, for the example in which I want to inform John that it is raining, it might be the case that a producer can more easily estimate the utility of 1 (e.g., if John is not looking). Therefore, the producer performs 1. It could equally be the case that a producer could more easily estimate the utility of 2 (e.g., the situation in which John is listening to music through headphones). Therefore, the producer performs 2. Finally, it might be the case that the effort required to estimate the utility of 1 vs 2 is too great, and therefore they perform 3. The two behaviours together are more likely to satisfy the prior intention than either on its own. In other words, the predictability heuristic assumes that as communication get harder, producers will produce both speech and gesture and not switch from one to the other, as predicted by the tradeoff hypothesis (De Ruiter, 2007). This view of the production of communicative behaviours is the predictability hypothesis.

To further emphasise the key insight of the predictability hypothesis, it is worth using an analogy from military training. There is a technique for close quarters combat known as the Mozambique drill, which states that in a kill or be killed situation shoot

a combatant twice in the chest and once in the head (in that order). The Mozambique drill has made its way into popular culture through various depictions in film and television. However, the underlying principle of the drill is based on the same logic as the predictability hypothesis. On the assumption that someone employing the Mozambique drill is in a kill or be killed scenario, their imperative is to make sure that they kill their combatant. Shooting someone in the head is a fairly sure way to kill them. However, shooting someone in the head requires precision that the time constraints of the situation will not allow. Shooting someone in the chest is less likely to kill them, but it requires less precision. Therefore, the Mozambique drill mimizes the effort required to make a single more deadly shot, by having the shooter perform three shots, two easy and one hard. Overall, performing all three shots is more likely to satisfy the intention of killing and not being killed. In this sense then, the choice to produce speech and gesture together is more effective at satisfying the prior intention an utterance producer had, even though the utility associated with performing multiple behaviours is lower.

One of the important features of predictability is that it does not inherently privilege speech or gesture and predictability at any one moment is the product of numerous factors. For example, knowledge of whether an interlocutor knows the meaning of the words you are using will have an effect. When speaking with people who do not share a language, it would be predictable that the words one uses will have a much lower expected utility than when speaking with those who do share a language. This may also be the case for particular lexis. The vocabulary of children's books is very different from the vocabulary of scientific journals. This suggests that knowledge that community members are expected to have (communal common ground (Clark, 1996)) affects predictability. Personal common ground (Clark, 1996), or knowledge shared between a two or more people, has a similar effect. This is evidenced by the literature on conceptual pacts and entrainment (Brennan and Clark, 1996), where shared knowledge regarding the use of a referring expressions increases the likelihood of that expression being used. Furthermore, it also increases the likelihood that a phonologically reduced version of word will be used. In other words, people are more likely to produce a reduced form when it is more predictable that their interlocutor will understand its meaning. There is

also clear evidence that predictability has an effect on gesture production as well. For example, if two people do not share a visual environment, then this negatively affects the expected utility of representational gestures. This is evidenced by the fact that people do not produce representation gestures when they cannot see each other (Bavelas and Healing, 2013). However, it has been demonstrated that not being visually co-present does not result in the complete removal of gesture. This makes sense if we consider that predictability must be partially determined by an individual's history as a communicator. Since gesture has been shown to increase the fluency of speaking for certain people, it seems likely that speech with gesture would be more predictable even when there is no one to see the gesture. In this sense, gesture can be a reflexive thing (it facilitates the production of actions) or a productive thing (it is directly geared towards satisfying a prior intention).

Where does this all lead? The general suggestion is that the more predictable the utility of an action is, the more likely an individual is to use that behaviour. This can be phrased in terms of communicative competence (cf. Bara, 2010). The better individuals are at accurately predicting behaviours with high utility the more communicatively competent they are. This description can be thought of as an intentional version of McNeill's "Mead's loop with a twist", which he describes in relation to mirror neurons (cf. Rizzolatti and Arbib, 1998) as:

Mead's loop refers to a posited new adaptation in the evolution of humans, wherein mirror neurons were "twisted" to respond to one's own gestures, as if they were from someone else

McNeill (2012, p. 64)

The point of McNeill's position is that in order to understand the significance of our own communicative behaviours, we must be able to perceive them as if we were an observer of that behaviour. It is this that predictability is acting upon. This leads to a critical feature of communication with gesture. If we only consider (expected) utility then the expected utility of two behaviours, for example, saying "John, it's raining" and the gestural behaviour of pointing out of the window is not additive, because it is a measure of



how well either behaviour satisfies a producer's prior intention. Therefore, the utility of two actions that both satisfy the prior intention will be less than the utility of either behaviour on its own, because of the added effort related to performing both actions. This leads to the suggestion that using speech and gesture to satisfy the same prior intention, will mean that one behaviour is classed as redundant. However, a large proportion of the literature discussed in this thesis shows that speech and gesture seem to convey the same information, even though at the semantic level unique information is conveyed through gesture. Much of the gesture literature focussing on production treats utility as the end product of communicative intentions. However, this "redundancy" is to be expected when the focus is shifted to predictability, because the predictability of two behaviours (even if they convey the same information) is greater than the predictability of either behaviour on its own. For example, if I say "John, it's raining" and point out of the window, both actions together are more likely to satisfy my prior intention because I am covering more bases without having to estimate utility.

Furthermore, predictability develops over time. In other words, the ability to predict which communicative behaviours are going to result in the highest utility develops differently for different individuals, depending on their other cognitive abilities and perhaps social conventions (cf. Kita, 2009). A range of interactions between cognition and gesture production were highlighted in section 2.2. For example, individuals with low phonemic fluency but high visualisation skills produce more gesture. According to the predictability hypothesis, this is because they are better at predicting the utility of gesture than the utility of speech (Hostetter and Alibali, 2007). Our ability to predict which action is most likely to satisfy our prior intention is of crucial importance, and therefore, it makes sense for our communicative competence to play to general cognitive strengths. This view explains the individual variability found in many gesture studies. Some people gesture a lot and others gesture hardly at all. Such variability has been explained using different models of how gesture is tied to cognition. While the predictability hypothesis acknowledges that the amount an individual will gesture is tied to their general cognitive abilities, it is because communication generally, and not just gesture, is tied to cognition. For example, some people may be adept at the sort of processing involved

in dealing with long, complex sentences. Such individuals will be expected to be able to better predict the utility of long, complex utterances. Therefore, they might be expected to produce long, complex utterances. However, the same person might not be very adept at spatial reasoning and, as a result, they may be less likely to gesture. In other words, predictability plays to our strengths.

This final point can help explain McNeill's two tribes theory. McNeill suggests that there are two groups of people, who have distinct psychological characteristics. McNeill (2015, p. 34) states that:

In the “[Growth Point]-tribe”, language and thought are non-linear, dailectical and unmodular; while in the “[Information Packaging Hypothesis]-tribe” they are linear, non-dialectical and modular

For McNeill then, there are two types of people, psychologically speaking. However, since the predictabilty hypothesis assumes that gesturing will develop alongside other cognitive abilities, it suggests that there are probably more than “two tribes”. The distinction between people is not simply due to their cognitive abilities, but to the way those abilities interact with the communicative practices.

This section has posited a new view of gesture production based on the notion of predictability. This view is based on key notions in pragmatics and decision theory and suggests that utterance producers do not base their decisions of what communicative behaviour to produce on the utility of any behaviour, but instead base their decisions on how predictable the utility of a behaviour is. Furthermore, because, in general, estimations of utility are practically intractable, utterance producers maxmimise predictability by performing multiple acts instead of one (highly utilituous) act. It is possible to explore the functionality of this position by reinterpreting an example from the map task.

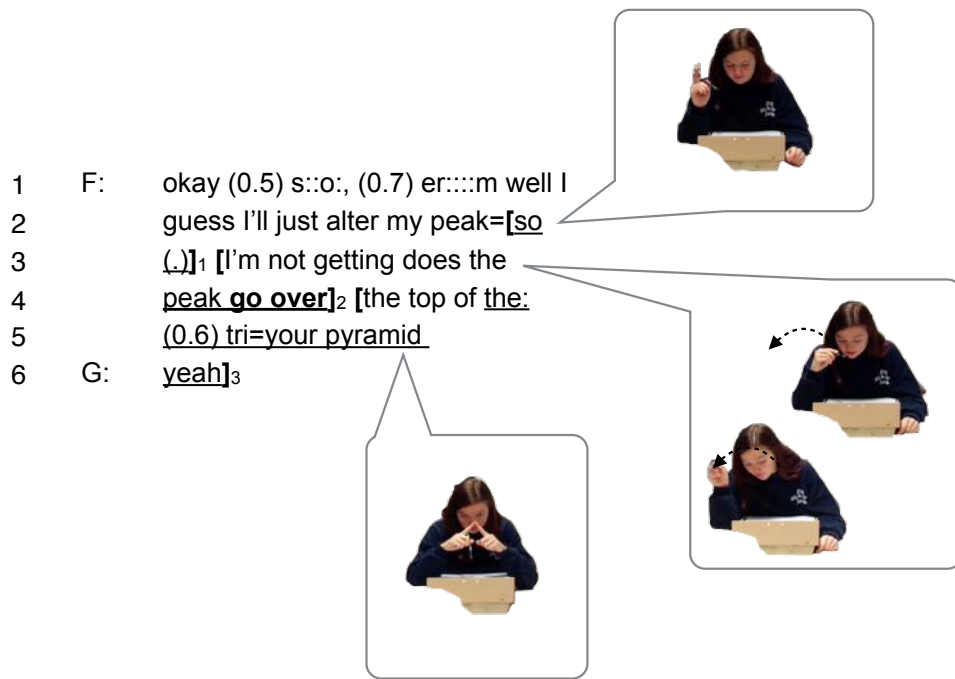


Figure 7.1: Example of Map Task Gesture

(a) Follower's Map



(b) Giver's Map

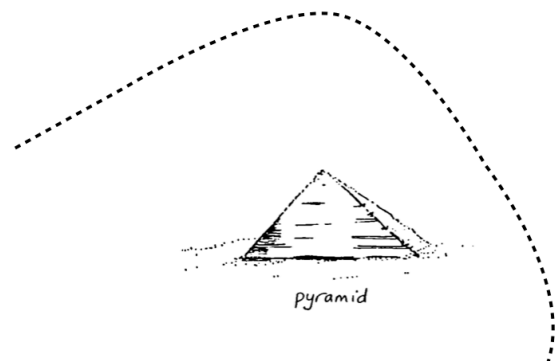


Figure 7.2: Section of maps being discussed in example 7.1

In example 7.1 F is asking G to clarify their description of the route as it goes over the pyramid (shown in figure 7.2). F produces two gestures. The first is a tracing gesture (lines 3 and 4) that depicts the shape of the route. The second gesture is a modelling gesture and depicts the triangular shape of a pyramid. This first gesture is co-referential with the phrase “go over” which, in conjunction with the word “peak” describes the shape of

the route. The gesture also depicts route shape. Although there are different semantic elements represented in the gesture (the curve of the route, for example) it largely conveys the same information as the speech. The second gesture models the triangular shape of the pyramid. Since F has an old temple on her map where G's pyramid is, F will have to retrieve the name of the pyramid from memory. This points to an important feature. Since her landmark is an old temple and she has not seen the pyramid on G's map, she does not know exactly what the pyramid looks like. Therefore, although the relationship between the route and the old temple and the route and the pyramid will be the same (i.e., the route will go over the pyramid / old temple), F may have difficulty conceptualising exactly what the route will look like on G's map. The phrase "go over" is vague in that it does not explicitly specify the shape of the route. If F and G shared the pyramid then the phrase could be anchored to that landmark. Therefore the fact that they do not both have the pyramid on their maps lowers F's ability to predict the utility of the spoken component (i.e., "go over") of the utterance. The gesture, however, specifically, depicts the shape of the route. Thus although speech and gesture seem to convey similar information, gesture is, in fact, increasing the overall predictability of the utterance, by providing a candidate for the unspecification of language. In terms of prior intentions and intentions-in-action, F's prior intention is to get G to clarify the route as it goes over the pyramid. In order to do this, F provides a candidate: "does the peak go over the top of the: (0.6) tri=your pyramid". In this case then F's prior intention is to offer a potential candidate for G to clarify his description. This prior intention produces two intentions-in-action. The first is to describe the route and the second is to gesture it. While both alone might be able to present the candidate clarification for G, together they are more likely to perform this function, and ultimately satisfy F's intention to get G to clarify their previous description.

### **A model for the pragmatic analysis of gesture**

Up to this point, this discussion has mainly been concerned with modelling the speech and gesture production from a theoretical perspective. Here, this model will be used to develop a unit of analysis for the pragmatic study of gesture. The model discussed

so far is in line with the three-part structure of signs (outlined in 2.2). Therefore, we can represent this theory using the on-switch, search, and off-switch taken from Enfield (2009b). This is visually represented in figure (7.3).

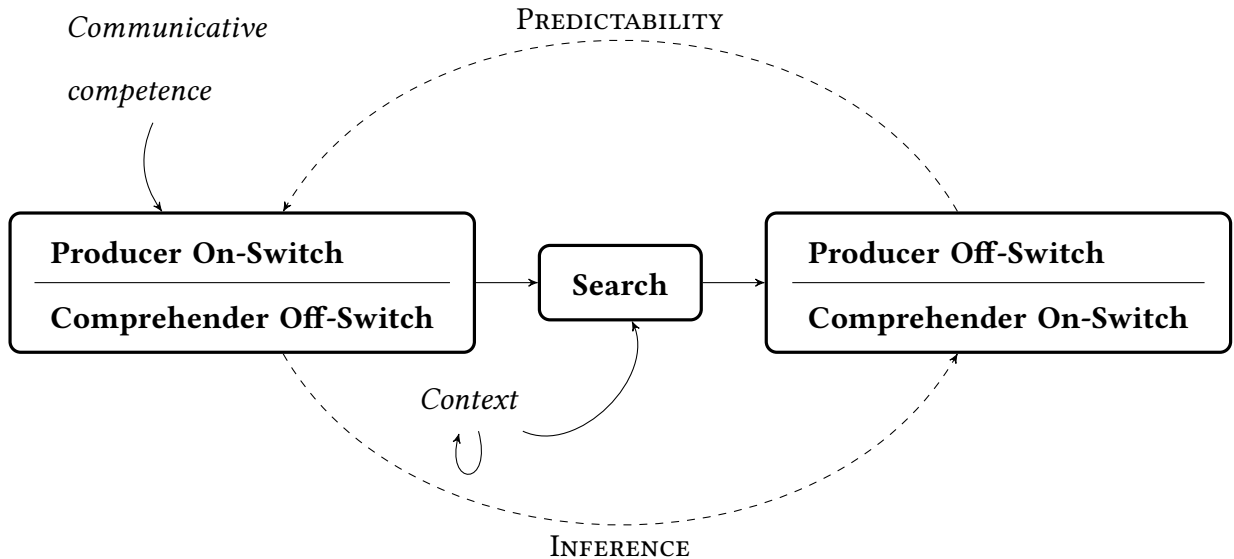


Figure 7.3: Model of Production and Comprehension (adapted from Enfield (2009b))

This model represents the process of production as a three stage process involving a Producer On-Switch (which is the producer’s prior intention), the search stage (which for the producer would involve selecting the intentions-in-action that will satisfy the producer’s prior intention) and finally the off-switch (which would be the production of the behaviour satisfying the intentions-in-action). Predictability is an inference from the behaviour to the prior intention and the production of that behaviour is what satisfies the prior intention. Finally, it is communicative competence that constrains the initial stage of the process. The model also depicts the stages a comprehender must go through in processing some behaviour. However, this time the on-switch and off-switch are reversed. It is the visible behaviour that starts the process, acting as the on-switch. The search stage involves narrowing down the behaviours that form part of the producer’s communicative intention from those that do not. And the off-switch is the realisation of the producer’s informative intention.

However, as outlined at several stages throughout this thesis, it is possible to represent meaning, not as mental states but as stages in an interactive sequence. Following a three-stage approach there is no real need to change the underlying principles of the

model. From a sequential perspective, the purpose of a communicative act is not something that is directly observable, but only inferable from how appropriately it is fitted to the previous action and whether or not the following action is expected. Thus, an utterance consisting of the word “Hello” is likely to appear as a greeting, which means that it is most appropriately fitted to the beginning of interactions. What’s more, the appropriate response to an utterance of “Hello” is another utterance of “Hello”. Therefore, from a sequential perspective the appropriateness of a behaviour is based on some prior behaviour or situation. Furthermore, this can be thought of from another perspective. An individual who produced “Hello” would do so expecting the next turn to be appropriately fitted to their turn. From this perspective, a behaviour’s success can be viewed as how well it sets up the context for the next turn to be appropriate. In other words, the meaning of an utterance is determined by its purpose, which is determined by both interlocutors together. Furthermore, the effect of each behaviour is in terms of how well it pushes forward the activity. Therefore, the model introduced above is not suited to this phenomena because it is not enough to explain the cognitive processes of individuals. It is necessary to also explain their role in the collective joint activity. According to downward evidence, a fitted next move is a representation of understanding at all levels on the action ladder.

Following this view, one might argue that since evidence of understanding is always in the uptake, then evidence of a project’s completion is always in the move immediately following the actual completion of the project. For this reason Bavelas et al. (2012) suggest that the fundamental unit of analysis should not be the projective pair itself, but the projective pair plus the next turn. This three part structure is referred to as a *grounding sequence* in which a participant presents a first pair part, which is taken up by the recipient in the production of a suitable second pair part, thereby displaying whether they have understood or not. However, the recipient cannot be sure if their second pair part is appropriate without some further response from the producer of the first pair part, which may acknowledge understanding. This third part of the sequence may be the initiation of a new project and thus only give indirect evidence that the uptake was an appropriate one.

With this in mind, we may adapt our model to look something more like this:

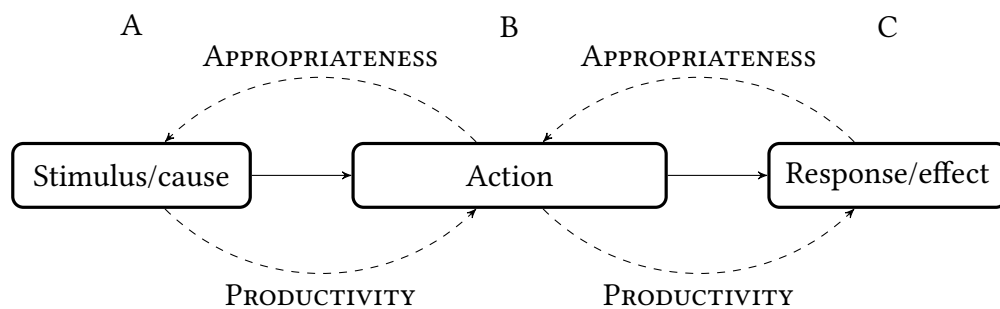


Figure 7.4: A model of communicative purpose. Adapted from Enfield (2009b)

The original model, represented in figure 7.3 above, consisted of three parts. The first was an on-switch, representing the stimulus that starts the process; the second was a search constrained by heuristics including the ground of the stimulus and context, and the third was an off-switch, representing the point at which some decision had been made. These three parts are cognitive phenomena and, as such, are not observable directly. However, the purpose of this adapted model is to capture the relationship between a behaviour and the sequence of events that led up to that behaviour, which are observable, and the future events that behaviour projects. It is related to the distinction made in Bara (2011, p. 453) between *communicative competence* and *interactional schemas*—what we are calling activity types. The former relates to the an individual’s ability to predict what some behaviour means, and the latter relates to an individual’s ability to predict what some behaviour is doing.

The argument I am pursuing is that there is no real need to change the theoretical underpinnings of the model, just our perspective on it. Almost every action produced by an agent is simultaneously a stimulus for further action and a response to some prior action, each productive in pushing forward the chain and appropriately fitted to the previous unit. Appropriateness in this sense is commensurate with the conversation analytic notion of *conditional relevance* (Schegloff, 1968); it is the non-cognitive version of predictability. It is also reflected in Austin’s *felicity* conditions on performative utterances (Austin, 1962) and Bara’s *validity* conditions on behavioural games (Bara, 2011, p.455). An appropriate next action is one that is predictable (or normative) given the current situation, however, as before, a purely predictable behaviour may not be productive

as a non-predictable one. Understanding the difference between communicative competence and how behaviours are fitted to activity types is of fundamental importance when developing a model of language use and what it is to have meaning.

However, in order to fully assess the notion of predictability it is important to compare it to current theories of gesture production. The rest of this section will be broken into three subsections, each focussing on a different model of production.

### 7.2.2 Predictability vs Tradeoff

The fundamental assumption of the tradeoff hypothesis (de Ruiter, Bangerter, and Dings, 2012b) is that when speaking becomes difficult, people gesture and when gesture becomes difficult, people speak. However, this model rests on the assumption that the goal of a communicative behaviour is to produce a signal with informational content. Therefore, gesture steps in when speech cannot produce that content and speech is used when gesture cannot. This model, therefore, suggests that our communication system is geared towards utility, based on the assumption that redundancy will be avoided. However, as I have been arguing, the communicative system is geared towards predictability and, as a result, “redundancy” is to be expected.

To my knowledge, there has been one major attempt to explore the suggestions of the tradeoff hypothesis, which can be found in de Ruiter, Bangerter, and Dings (2012a). This paper largely fails to find evidence in favour of the tradeoff hypothesis and the authors state that their results favour the “hand-in-hand” hypothesis (So, Kita, and Goldin-Meadow, 2009). The experiment discussed in de Ruiter, Bangerter, and Dings (2012a) involved two participants, who could both see a collection of tangrams on a wall in front of them. They are sat side-by-side. One of the participants, the director, is presented with a tangram on a computer screen and they must direct the other participant, the matcher, to that object so they can identify it. de Ruiter, Bangerter, and Dings (2012a) manipulated three elements designed to affect the difficulty/ease of speech production. First, codability was manipulated by having tangrams of differing complexity. This is on three levels: simple (shapes that can be described with a single word, such as “star”) humanoid (shapes that look like human characters); and abstract (shapes that have no obvious char-



acteristics). The next thing they manipulated was common ground. Common ground was manipulated by repeating the last three trials for each codability condition twice. Therefore, directors were required to describe 24 items (8 in each codability condition) followed by a further 18 items (3 in each codability condition, appearing twice). This meant that those items were described three times, allowing for them to be grounded. The last manipulation was visibility, which was manipulated by putting a barrier in between the two participants. de Ruiter, Bangerter, and Dings (2012a) used the frequency of gesture per hundred words their dependent variable. Gestures were broken into three categories: pointing gestures, obligatory iconic gesture (which present information not in speech), and nonobligatory iconic gestures (which present information also found in speech). Therefore, according to the tradeoff hypothesis, de Ruiter, Bangerter, and Dings (2012a) predict that gesture will increase as codability increases, as common ground is reduced, and should not be produced when participants cannot see each other.

The main findings of de Ruiter, Bangerter, and Dings (2012a) are shown in figure 7.5.

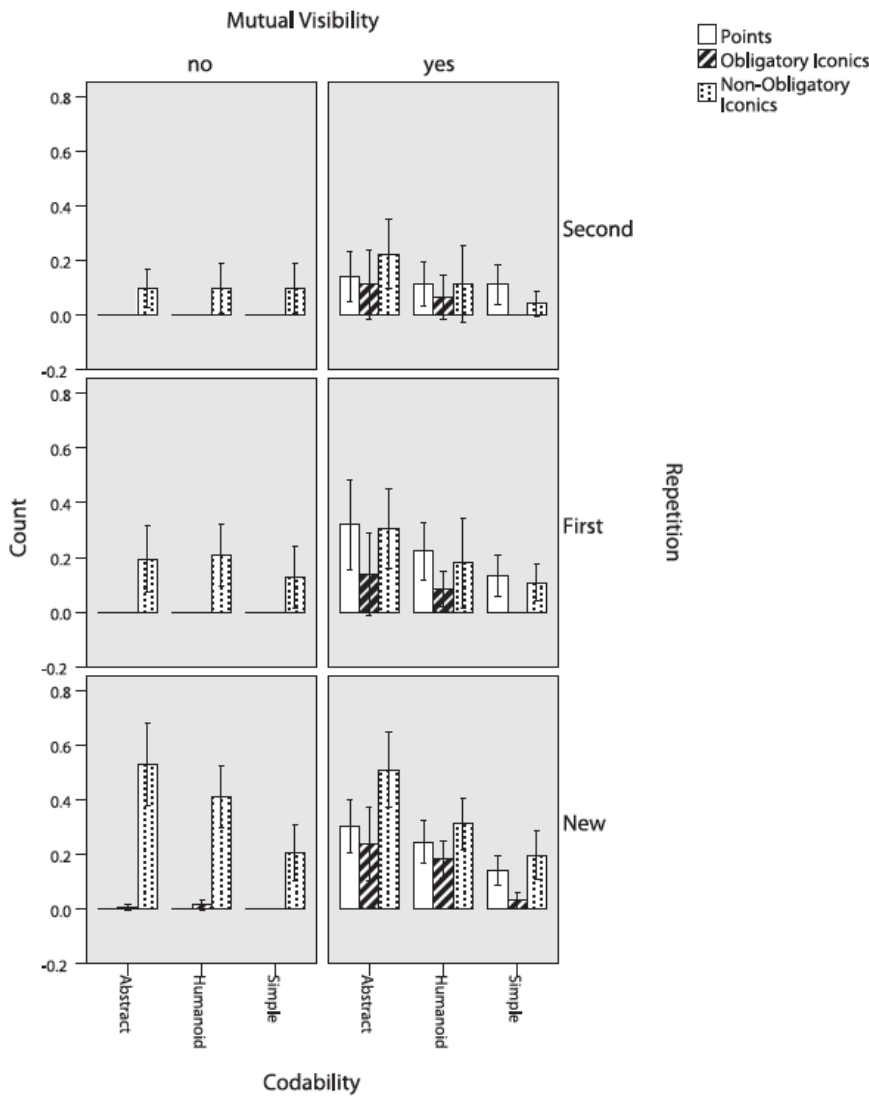


Figure 7.5: Results for gesture rate from de Ruiter, Bangerter, and Dings (2012a)

### Results for speech

de Ruiter, Bangerter, and Dings (2012a, p. 239) found that repetition reduces the number of words used. This is entirely consistent with the literature on grounding (cf. Clark and Brennan, 1991; Mushin et al., 2003; Bavelas et al., 2011). It was also found that higher codability led to a reduction in the number of words. Additionally, they found that mutually visible pairs produced fewer words than hidden pairs. This final finding is not fully developed but could be explained using the notion of joint attention (Tomasello, 2010) or perceptual co-presence (Clark, 1996). In the mutual visibility condition, it is possible to see whether an interlocutor is attending to the screen containing the tangrams, therefore, a director would be better able to predict the utility of their referring expression. In

this sense, if we assume that less predictability may result in longer descriptions, then it makes sense that mutual visibility will reduce the number of words in speech.

### **Results for pointing gestures**

Turning to their results for gesture, they found that directors did not produce pointing behaviours when they could not see their partner. They interpret this as suggesting that pointing gestures are communicative. They found no effect of codability or repetition on pointing gestures. In other words, the difficulty of encoding an item did not affect the rate of pointing gestures. Nor did people produce fewer pointing gestures on repeated references to the same items. This, they argue, is against the tradeoff hypothesis. However, de Ruiter, Bangerter, and Dings (2012a) do actually demonstrate an increase in pointing gesture rates. Moreover, it was shown that pointing gestures occur more with locative descriptions and decrease with conceptual pacts. The fact that gesture does not decrease with repetition but do reduce with the establishment of conceptual pacts seems to suggest that these two things are independent. This is unusual since conceptual pacts are commonly found in repeated reference. Therefore, it is possible that the participants are adopting two strategies for repetition. One is to reproduce the repetition (including gesture) and the other is to establish a conceptual pact and not gesture. There is a potentially clear explanation for these results. Describing the referent gets harder as codability decreases. This is because encoders have to rely less on lexical items (e.g., star) or on descriptions of similarity (e.g., like an ice skater) and more on ad hoc spatial descriptions. As spatial descriptions become more complex, the ability of a producer to predict the utility of their utterance will be reduced. Therefore, producers may look for alternative strategies. One such strategy would be to use locative descriptions coupled with underspecific referent descriptions, for example “the triangular shape in the top right corner”. However, in English, which employs a relative frame of reference, locative descriptions tend to be produced from a producer’s perspective (Levinson, 2003) and it is this perspective-specific nature of locatives that results in them being coupled with pointing gestures. This was found to be the case in the map task, where utterances produced from the producers perspective (e.g., left/right terms) were more likely to be

accompanied by gesture. The relationship between left/right terms and gesture has also been established in the literature (Melinger and Kita, 2007).

Moreover, pointing gestures are not spatial in the sense that they depict spatial features of a referent, but are designed to direct comprehenders to certain area of visual space, constraining the comprehender's field of vision (cf. Clark, 2003). Therefore, pointing gestures do go hand-in-hand with locative descriptions but they do not express the same information. The results of de Ruiter, Bangerter, and Dings (2012a) are consistent with a situation in which participants produce fewer referent descriptions and more locative descriptions as the codability of the referent reduces. It is possible that directors are adopting two referring strategies, one where descriptions, lexical item, or entrained referring expressions are used to direct the matcher to the referent and another where a locative description plus pointing gestures are used. However, these suggestions are not tied to any particular situation but to the particular participants' preferences. These results are actually in line with the findings of the map task (chapter 4) where it was shown that participants adopted two strategies. In one they refer to the landmark directly and do not produce as much gesture and in the other they described the route producing tracing gestures. Gesture production is not necessarily tied to the ease of speaking, but the strategy adopted by the producer.

### **Results for obligatory iconic gestures**

Like pointing gestures, obligatory iconics were not produced when participants could not see each other. de Ruiter, Bangerter, and Dings (2012a) interpret this as meaning that obligatory iconic gestures are produced to be communicative. de Ruiter, Bangerter, and Dings (2012a) report that there is no affect of repetition or codability on the rate of obligatory iconics. In terms of codability, their results show extremely large error bars for the rate of obligatory iconic gestures, the error bars for first and second repetitions are more than twice the average. This suggests a lot of variability in the results. Obligatory iconics, as de Ruiter, Bangerter, and Dings (2012a) describe it, require explicit reference to the gesture in speech. It is possible that this kind of gesturing is susceptible to individual variation. Moreover, it was shown in the map task that obligatory ges-

tures (coded as speech framed) were associated with increased agency. In other words, the biggest distinction between obligatory and non-obligatory iconic gesture is overt control over the gestures. It seems, then, that the most obvious explanation for the findings of obligatory gesture rates is individual variation. However, it is also possible that the rate of obligatory iconic gestures remains constant because speech and gesture both increase or decrease. That is, as gesture is reduced, speech is reduced and as gesture increases, speech increases. This is further emphasised by the finding that the number of features described increases with increased gesture rate.

### **Results for nonobligatory iconic gestures**

One of the most important gesture types for the theory progressed here are what de Ruiter, Bangerter, and Dings (2012a) call nonobligatory gestures. However, it is not clear exactly how these gestures are tagged. The example they describe as nonobligatory involves a participant saying “the big triangle” while tracing a triangle. However, to what extent can this really be considered nonobligatory? It is likely that the gesture would depict the orientation of the triangle, depicting which point was at the top. It is also possible that the gesture would depict what type of triangle was being described. In Watson and Wilson (Submitted) it is demonstrated that different gestures are produced when individuals are referring to isosceles triangle compared to right-angled triangles. Additionally, Cohen, Beattie, and Shovelton (2011) show that when explored at a semantic feature level, over 80% of gestures convey unique semantic information. It is for this reason that a semantic feature analysis, such as the one presented with the map task above is crucial for analysing the relative contribution of speech and gesture. However, de Ruiter et al.’s results are still of interest. They show that nonobligatory iconics are not affected by visibility. de Ruiter, Bangerter, and Dings (2012a) interpret this as meaning that nonobligatory iconics are not produced for communicative purposes. However, since they do not report it, it is impossible to know whether there is an effect of semantic features that were uniquely presented through gesture (future research could try and replicate this result using a semantic feature analysis). Nor was there an effect of repetition or codability on the rate of nonobligatory iconics. As with obligatory iconics,

the number of features described increased with gesture rate. In other words, directors produced a greater proportion of nonobligatory iconic gestures when they were talking in more detail about the features of the tangrams. Therefore, it is possible that there is an effect of repetition and codability on nonobligatory gesture, but this is being masked by an simultaneous increase in speech.

### Summary

Overall, de Ruiter, Bangerter, and Dings (2012a) suggest that their study found little evidence for the tradeoff hypothesis. The biggest thing to note is that there was a great deal of variation. Simply looking at the average scores shown in figure 7.5, shows that the data is in line with their hypotheses. One answer to the amount of variability is that the hypothesis that people either do gesture or do not gesture is far too reductive. One of the suggestions of the predictability hypothesis is that, because predictability is based on communicative competence, there is a great deal of variability in the strategies people adopt. Therefore, it is entirely possible that some people will adopt strategies consistent with the tradeoff hypothesis. By focussing on only the presence/absence of certain types of gesture, it is likely that those findings that do go in favour of the tradeoff hypothesis will be hidden by alternative strategies. de Ruiter, Bangerter, and Dings (2012a) do show that pointing increases with locative descriptions and is reduced with conceptual pacts. They also show that obligatory and nonobligatory gestures increase with additional spatial features. However, because there is no explicit reference to locative descriptions, conceptual pacts or the number of features described in speech in relation to the different conditions, it is difficult to interpret what effect these features have on the overall findings. However, it is possible to interpret all of these findings as being in line with the predictability hypothesis. First, in terms of locative descriptions, de Ruiter, Bangerter, and Dings (2012a) define two types of locative descriptions: absolute (e.g., “the upper left corner”) and relative (e.g., “below the big triangle”). This example of an absolute locative include spatial terms that require a comprehender to adopt a producer’s perspective. In the map task, reported above, it was shown that participants gesture about the direction of the route when the speech includes egocentric spatial descriptions. This

is also consistent with research that shows that left/right spatial descriptions are commonly accompanied by gesture (Melinger and Kita, 2007). In relation to the predictability hypothesis, it is argued that because left/right spatial descriptions typically require comprehenders to adopt the producer's perspective, the ability to predict the utility of the utterance is reduced, on the assumption that processing a description from an interlocutor's perspective takes more effort. As a result, producers use gesture to raise the predictability of the overall utterance. This is why absolute locative descriptions are accompanied by pointing gestures. Relative locative descriptions are underspecific for different a different reason. The description "below the big triangle" could refer to any item beneath the big triangle and not necessarily the object immediately under. Therefore, once again, the underspecificity of the spatial description lowers predictability and therefore gesture raises it. Although de Ruiter, Bangerter, and Dings (2012a) do not report it, the predictability hypothesis would predict that directors would use different locative descriptions in the condition where participants are not mutually visible.

Second, the establishment of conceptual pacts are associated with a drop in gesture. This is entirely consistent with the predictability hypothesis. Conceptual pacts occur when communicators entrain upon the use of a lexical item or referring expression, such that the expression become ad hoc names for the object to which it is being used to refer. The classic example from Clark and Wilkes-Gibbs (1986), which shows the subsequent use of referring expressions, all describing the same object exemplifies this process:

- (7.5)
- All right, the next one looks like a person who's ice skating, except they're sticking two arms out in front.
  - Um, the next one's the person ice skating that has two arms?
  - The fourth one is the person ice skating, with two arms.
  - The next one's the ice skater.
  - The fourth one's the ice skater.
  - The ice skater.

In this example, the expression "the ice skater" starts as a description and ends up being a naming expression. According to the predictability hypothesis, as conceptual

pacts are formed the ability to predict the utility of the entrained expression is increased. Therefore, as conceptual pacts grow, gestures used to refer to the object should be reduced.

Third, an increased number of spatial features result in an increase in both obligatory and nonobligatory gestures. This was also shown in the map task. Since it is assumed that all utterances are subject to the predictability hypothesis, then spatial descriptions are most likely to be accompanied spatial gestures. What's more, it is not expected that spatial gestures will necessarily convey unique information, but that spatial gesture should convey information that is less predictably encoded in the spatial description. In other words, those features that are underspecific or indexical should be more likely to be accompanied by gesture.

It seems then that the notion of predictability is a better model for analysing gesture than tradeoff. However, de Ruiter et al.'s paper was written in response to a paper by So, Kita, and Goldin-Meadow (2009) that focusses on the hand-in-hand hypothesis. Therefore, the next section focusses on reinterpreting the results of So, Kita, and Goldin-Meadow (2009) from the perspective of predictability.

### 7.2.3 Predictability and the Hand-in-Hand hypothesis

The hand-in-hand hypothesis is an elaboration of the interface hypothesis and suggests that gesture mirrors speech. The question this section asks is whether predictability can explain the findings of So, Kita, and Goldin-Meadow (2009). They showed that when recounting stories that included either two males (M-M) or a male and female (M-F), participants were more likely to use pronominal reference in the M-F condition than the M-M condition. Participants tended to use noun phrases (e.g., “the one with the noose”) instead of pronouns in the M-M condition. The noun phrases used in the M-M condition were less specific (more ambiguous) than the pronouns used in the M-F condition. This is because pronouns are gendered in English, and, as a result, are more specific in the M-F condition than the M-M condition. So, Kita, and Goldin-Meadow (2009) explored the use of gestures that placed a referent in space (similar to placement in sign language), where, for example, an producer might point to the right when they



refer to one character and to the left when they refer to another. So, Kita, and Goldin-Meadow (2009) were interested in whether such gestures were more likely to be used when referring expressions were ambiguous or specific. Results showing that gestures are more likely to be used with specific referring expressions would provide evidence for the hand-in-hand hypothesis, whereas evidence that gestures occur with less specific referring expression would suggest that gestures bolster the reference.

So, Kita, and Goldin-Meadow (2009) found that gestures were more likely to be used when a description uniquely specified a referent than one that was more ambiguous. They use this finding to suggest that gestures go hand-in-hand with speech, so utterances that more specifically refer to a referent are more likely to be accompanied by gesture that also specifies the referents. There is a different reading of these results. First, it is interesting that participants use more noun phrases than pronouns in the M-M condition than in the F-M condition. This suggests that noun phrases are being used because participants are not able to refer to a referent using a pronoun (because it would be completely ambiguous), however, when a pronoun does uniquely specify a referent, they use it. This is consistent with the predictability hypothesis. A participant is better able to predict that their interlocutor will be able to interpret a referent with a noun phrase when a pronoun would not uniquely specify it. When a pronoun does uniquely specify it, that pronoun would be the easiest way to guarantee the utility of the referring expression. Therefore, the central question is why are participants more likely to produce gestures with uniquely specifying referring expressions? Crucially, pronouns make up a lot of the data for this analysis, so it is pronouns that uniquely specify referents that are largely responsible for the finding. In other words, it is possible that people gesture more when they produce pronouns than when they produce noun phrases. Gestures that place referents in space are similar to pronominal placement in sign languages. Pronominal placement allows sign language users to refer to multiple referents in space over time (in the sign language literature such pointing behaviours are often referred to as indexes (cf. Sutton-Spence and Woll, 1999; Johnston and Schembri, 2007). English pronouns, on the other hand, only allow individuals to refer to a maximum of two individuals uniquely and only in the condition where one

is male and the other is female. Therefore, while these pronouns are more specific in this experiment, they generally are not<sup>1</sup>. The noun phrases used in the M-M condition would uniquely specify the referents even if a new individual was introduced later in the narrative, whereas the pronouns would not. It is possible that placing referents in space gesturally is a means of keeping track of pronominal referents because they are not uniquely specified in the long term. For example, if Anne and Bob have established during a conversation that “he” refers to “David”, it would be unlikely that in a year’s time they could use “he” to refer to “David” without first re-establishing it. However, utterances such as “Do you remember the guy who served chips at primary school” are not uncommon. Noun phrases have referential power over the long term, whereas pronouns are a relatively short term solution. This could be explored by introducing a third manipulation where a new character is introduced into the narratives. In the M-F condition, it is not clear how the speakers would deal with the (new) inability to uniquely specify referents with pronouns. For example, in a narrative with M-F where a new F or M were introduced after the use of “he” to refer to the male character and “she” to refer to the female character had been established, it is not clear what the hand-in-hand hypothesis would predict the speaker would do. The predictability hypothesis suggests that the introduction of a new character would lower the predictability of referring pronominal referring expressions. Therefore, gestures should be used more when a new character is introduced than before. In contrast, what would happen in the M-M condition? Since a new character would not necessarily disrupt the noun phrases used to refer to the character in the M-M condition, it is predicted that there should not be an increase in gesture use. The hand-in-hand hypothesis would not be able to explain the changing need for more or less specificity in referring expressions. However, since the predictability hypothesis assumes that judgements regarding the type of utterance a producer is going to make are governed by predictability, then as the ability to estimate the utility of an utterance is reduced the more likely people will be to use gesture.

The hand-in-hand hypothesis explains that people often produce content in gesture

---

<sup>1</sup>Furthermore, although the noun phrases were more ambiguous than the pronouns in this experiment, it was still the case that comprehenders were able to interpret their referent 84% of the time compared to 91% of the time for pronouns. Therefore, although they were more ambiguous, the noun phrases were not ambiguous in a general sense.

that has a similar meaning to the content of speech. This hypothesis helps explain why more specific referring expressions were accompanied by gesture (which further specifies the referent). However, the hand-in-hand hypothesis cannot explain why participants opted to use noun phrases when pronominal reference would not have specified the referent. In other words, it is the use of pronouns and not specificity that results in gesture. The predictability hypothesis suggests that pronouns are accompanied by gesture because of the necessity to keep track of multiple referents when pronouns are used.

The motivation behind the hand-in-hand hypothesis comes from the interface hypothesis (Kita and Özyürek, 2003). Therefore, the next section compares the interface hypothesis to the predictability hypothesis.

#### **7.2.4 Predictability and the interface hypothesis**

The interface hypothesis (Kita and Özyürek, 2003) predicts that the content of gesture mirrors the content of the speech because it emerges in the interface between spatio-motoric representations the preparation of linguistic units for speech. Therefore, gesture will be shaped by speech and the spatio-motoric properties of referents. It is for this reason that gesture mirrors speech, but can also depict information that is not part of the linguistic representation. There is great deal of cross-linguistic (Özyürek et al., 2005) and language-specific evidence (Kita et al., 2007) for the interface hypothesis. However, while this perspective accurately captures much of what has been shown regarding the actual behaviours people produce, it does not do so from a pragmatic perspective. Instead, the interface hypothesis focusses on the cognitive aspects underlying the production of speech and gesture and not necessarily the communicative ones (see section 2.2, for a more detailed explanation). Is it possible to retain the findings of the information packaging hypothesis without violating the predictions of the predictability hypothesis? The view here is that information packaging can be subsumed under predictability. The central question posed by the predictability hypothesis is what motivates producers of communicative behaviours to produce the behaviour they did rather than another behaviour that might be able to achieve the same effect. The answer is that individu-

als produce the behaviour that most predictably achieves that effect. Prediction is not based purely on expected utility, but on previous uses of communicative behaviours. In other words, if a behaviour was successful before, it is more likely to be successful again. Predictability is not simply a feature of one-off communicative events, but is based on a lifetime of communicating. Therefore, gesture production should be language specific. What's more, because speech does not emerge in sentence length units, but chunks (sometimes called processing units) then predictability will be assessed at the level of the chunk and not the sentence. Therefore, if a chunk includes a spatial description representing conflated MANNER and PATH information and that aspect of the utterance is less predictable than the equivalent chunk with gesture depicting the same information, then a speech with gesture chunk will be produced. However, if a language does not conflate MANNER and PATH then the predictability of the MANNER chunk and the predictability of the PATH chunk will be assessed separately. Therefore, if either or both are less predictable than the equivalent speech only chunks then gestures will be produced together with their associated chunks.

In other words, conflated MANNER and PATH in speech will be accompanied by conflated MANNER and PATH gestures and decomposed MANNER and PATH speech will be accompanied by decomposed MANNER and PATH gestures. However, there are examples of MANNER and PATH gesture accompanying MANNER or PATH only speech. For Kita and Özyürek (2003) this finding is due to the fact that gesture is derived from spatial-motoric representations whereas speech is derived from linguistic representations. Spatial-motoric representation are not compositional in the same way that linguistic representations are, because spatial representations represent objects holistically. Therefore, gestures may present more information than speech. This is entirely consistent with the predictability hypothesis.

This can be explained with the results of the “around” analysis in the map task. These findings are exemplified with the following example:

(7.6) “and then do a loop around again”

The word “around” was coded as including MANNER information because it conveys information about the shape of the route. However, here “around” is underspecific be-

cause it does not convey information about the DIRECTION of the route, is it not clear whether it is travelling leftwards or rightwards. This utterance was accompanied by a tracing gesture that depicted both the DIRECTION and MANNER of the route. This suggests that gestures are used to increase the predictability of route descriptions including the word around. Furthermore, it was shown that in utterances that described the DIRECTION of the route, but did not include gesture were more likely to include reference to a landmark, which was used to anchor the route. In this case, although the producer is describing direction they are not using gesture because the common ground associated with the landmark increases the predictability of the route description. Therefore, while Kita’s model correctly predicts the distribution of speech and gesture, it does not necessarily explain why factors such as common ground have an effect on gesture production. Predictability, since it is not a feature of speech or gesture but a feature of communicative behaviours, generally explains why people gesture in certain contexts and not others.

### 7.2.5 Summary

The model of speech production hypothesised in this chapter can be summarised using Searle’s representation of the structure of actions.

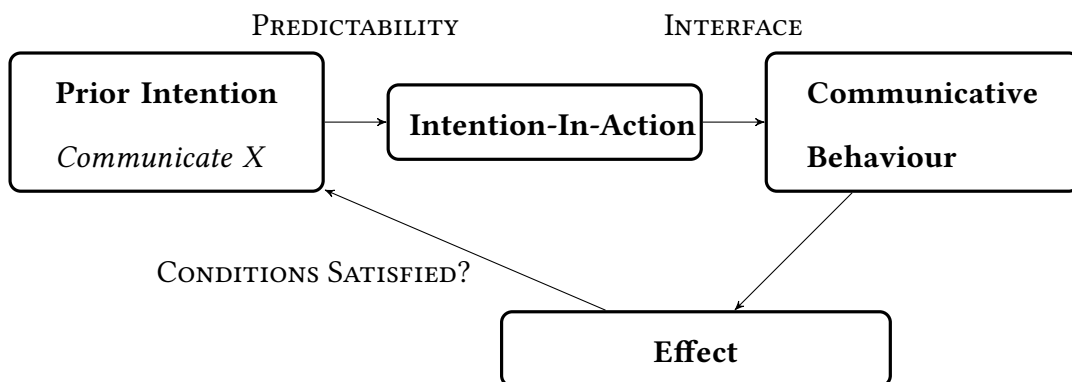


Figure 7.6: Model of Speech and gesture production

This model suggests that predictability governs the choice of whether or not an individual will produce gesture. The process described in the interface hypothesis governs the synchronisation of speech and gesture. However, a prior intention is not satisfied once a communicative behaviour has been produced but once its effect is observed. If the prior intention is not satisfied then the prior intention will either be reproduced or

abandoned. Responses that do not satisfy prior intentions will lower the predictability of the behaviour that was produced the first time around. This is a crucial insight of this theory because it suggests that unsuccessful reference will result in a change in behaviour. This is potentially highlighted by the effect on the projects in the map task. Gesture was less likely to occur in more embedded projects. In other words, it is expected that unsuccessful communicative behaviours containing gesture might result in reformulations that do not include gesture.

### 7.3 Limitations and Future Directions

This section outlines the future directions of the studies reported in this thesis, suggesting potential shortcomings.

#### 7.3.1 Map Task

The map task is an excellent paradigm for exploring the intentional use of gesture, because if the analyst has access to informational content (i.e., the route) that the giver is trying to describe. Therefore, the analyst is in a position to explore the various ways in which the participants achieve their goals of jointly coming to an agreement regarding the route. The results discussed above (section 4.2.9) were compared to corpus studies focussing on speech and gesture composition (e.g., Mehler, Lücking, and Weiss, 2010; Lücking et al., 2013). While such comparisons are warranted, the smaller dataset associated with this thesis increase the effect of individual variability. Therefore, an obvious next step for the map task is build a corpus to explore the validity of the findings reported here. Additionally, while the results reported in this thesis focus on the semantic features distributed across speech and gesture, projects were only explored generally. Therefore, the findings could be further explored at the level of individual sequences. For example, this could be achieved by investigating whether or not an utterance is followed by an affirmative or negative utterance and how the initial speaker reformulates their original utterance. Then the formulation of the original utterance could be compared to the reformulation. According to the principles of the predictability hypothesis, the first utterance the producer produced would have been the one that would most pre-

dictably satisfy their informative intention. Therefore, the fact that it was not successful means that this was not the case and their reformulation would be based on a new prediction. By tracking the types of utterances that occur throughout a map task and the effect negative responses have on them, one could assess the validity of the predictability hypothesis. According to this line of argument, it could be expected that participants will not reuse strategies that have already been demonstrated not to work.

The map task has also recently been used to explore the composition of speech and gesture in endangered Modern South Arabian Languages. The results are reported in (Watson and Wilson, Submitted) and they provide a further demonstration of the interface hypothesis. For example, it was shown that languages which tend to decompose direction and orientation information (e.g., “the route goes down the side of the pelicans on the right”) compared to those that conflate direction and orientation (e.g., “the route goes down the right of the pelicans”), produce different gestures. Speakers of languages that decompose direction and orientation tend to produce two gestures, one depicting the direction and the other depicting the orientation. However, speakers of languages that conflate path and orientation tend to produce a single gesture depicting the direction, but do not produce orientation gestures. The project aims to explore this finding further.

### 7.3.2 Visual World Study

There have been very few attempts to use the visual world paradigm to explore the effect of gesture on real-time comprehension of utterance. The visual world paradigm reported in chapter 6 demonstrates that it is an excellent resource for exploring, not just initial effects of speech and gesture but the ongoing effect of speech and gesture composition on the comprehension of utterances. The stimuli used in the study are problematic for two reasons. First, the gestures are all produced from the perspective of the comprehender, rather than producer. Because gestures tend to be produced from a producer’s perspective, it is critical to fully validate the findings that the study is conducted with additional gestures produced from the producer’s perspective. This would have the advantage of showing that the findings of the eye tracking study can be generalised to more typical

gestures.

The second potential problem rests in the gestures used in the condition where only speech conveys MANNER information. It's possible that because the flat hand gesture used in this condition could be interpreted as suggestions that the route depicted is flat, that the gesture is judged as conveying incorrect information rather than only DIRECTION and ORIENTATION information. One way to explore this would be to investigate if there is a difference between a video that does not include gesture at all and the speech only and neither speech or gesture conditions. If the gestures without MANNER are misleading participants, then the no gesture condition would not be expected to result the late effect that was demonstrated above. In other words, the no gesture condition should show similar result to the conditions where MANNER is conveyed in gesture.



# Chapter 8

## Conclusions

The purpose of this thesis was to explore gesture from a pragmatic perspective. A pragmatic perspective must be able to explain two key features of communication. First, it must be able to explain how/why utterance comprehenders take certain behaviours to be communicative. Second, it must be able to explain why an utterance producer produced the behaviour they did and not some other one.

Communicative acts are embedded within activities and it has been shown that gestures are tailored to the needs of the activity. Gestures were more likely to occur when producers are describing task-critical information and are less likely to occur when confusion arises. In other words, from a top down perspective (from activity to act), gestures are a crucial resource that people use to progress the activity. Gesture is not random.

However, a theory of activity is not a theory of pragmatics. This is because while behaviours may have a demonstrable effect on the activity, it is impossible to explore this effect at the level of production and comprehension. While it is possible to say that a certain behaviour had a certain effect, it is not possible to say that an utterance producer produced a certain behaviour to have a certain effect. Nor is it possible to say why a comprehender interpreted that behaviour in the way they did. Therefore, it is crucial to understand why behaviours take the form they do and what effect the form of behaviours has on comprehension. It has been shown that the semantic content of speech and gestures is determined by what an utterance producer wants to convey. It was also shown that gesture effects comprehension by increasing the comprehender's confidence in what producer is trying to convey. In other words, gesture has a demonstrable effect

on both production and comprehension.

In terms of comprehension, these findings are largely in line with Enfield's model of sign filtration (Enfield, 2009b; Enfield, 2013). What is more, the findings additionally provide an explanation for why Enfield's heuristics are useful for comprehenders. Paying attention to gesture does not simply provide a comprehender with an additional information resource, it helps comprehenders build stronger representations of what producers are trying to communicate.

However, to my knowledge, this thesis represents the first theory of gesture production that is based on the principles of pragmatics. That is not to say that theories of gesture production do not exist. However, they tend to treat utterance production as a process that aims to create an informationally complete utterance or they focus on the effort required to produce utterances. Here, it has been suggested that effort is a poor measure for the principles that guide utterance production. This is because it is difficult as an analyst to speculate about the effort involved in choosing one particular behaviour over another. Therefore, it seems unlikely that the largely unconscious process of utterance selection is based on a calculation of effort. The alternative is that participants base their choice of utterance on how predictable it is that that utterance will have the producer's intended effect. However, while utterance producers may not be able to calculate the effort an utterance might take, they will certainly be able to predict the effect it will have. From this point of view, overdoing it is better. Presenting analogous information in both speech and gesture is more likely to get the utterance producer's message across than producing either element on its own. This position is referred to as the predictability hypothesis. If predictability guides production, then it may play a role in comprehension. Comprehension is a process of attributing intentions and if predictability plays a role in intentional utterance production, then it is likely that it is part of the inference a comprehender makes. Therefore, the two sides of production and comprehension can be represented in table 8.1.

PREDICTABILITY	
<i>Production</i>	<i>Comprehension</i>
Interface	Contextual association heuristic
	Unified utterance-meaning heuristic

Table 8.1: Principles of Production and Comprehension

Table 8.1 aims to represent how production and comprehension are mirrored. The general idea is that production is guided by a notion of predictability, which guides a producer to produce a certain type of behaviour. The interface hypothesis governs the unification of speech and gesture. Mirroring Enfield’s heuristics, the interface could be split into two heuristics. First, *maximise unification heuristic*, which states “if two elements of a composite sign can be used to refer to a single referent, produce both”. And second, *maximise co-temporality and co-proximality heuristic*, which states: “if two elements of a composite sign can be used to refer to a single referent, produce both of them to be co-temporal and co-proximal”. These heuristics produce behaviours that are picked up by Enfield’s comprehension process.

It is important to stress that interface is governed by predictability. So in other words, if a behaviour would more predictably satisfy the producer’s informative intention, but would violate the assumptions of the interface hypothesis, then that behaviour will be performed. This captures the fact that some people do not produce gesture, because predictability is satisfied by speech alone.



# Bibliography

- Alibali, M. W. and S. Kita (2010). "Gesture highlights perceptually present information for speakers". In: *Gesture* 10.1, pp. 3–28.
- Alibali, M. W., S. Kita, and A. J. Young (2000). "Gesture and the process of speech production: We think, therefore we gesture". In: *Language and cognitive processes* 15.6, pp. 593–613.
- Allopenna, P. D., J. S. Magnuson, and M. K. Tanenhaus (1998). "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models". In: *Journal of Memory and Language* 38.41, pp. 419–439.
- Altmann, G. T. (2004). "Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'". In: *Cognition* 93.2, B79–B87.
- Altmann, G. T. and Y. Kamide (1999). "Incremental interpretation at verbs: Restricting the domain of subsequent reference". In: *Cognition* 73.3, pp. 247–264.
- (2007). "The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing". In: *Journal of Memory and Language* 57.4, pp. 502–518.
- (2009). "Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation". In: *Cognition* 111.1, pp. 55–71.
- Altmann, G. and J. Mirković (2009). "Incrementality and prediction in human sentence processing". In: *Cognitive Science* 33.4, pp. 583–609.
- Anderson, A. H. (2006). "Achieving understanding in face-to-face and video-mediated multiparty interactions". In: *Discourse Processes* 41.3, pp. 251–287.
- Anderson, A. H. et al. (1991). "The HCRC map task corpus". In: *Language and speech* 34.4, pp. 351–366.
- Armstrong, D. F., W. C. Stokoe, and S. E. Wilcox (1995). *Gesture and the Nature of Language*. Cambridge University Press.
- Atlas, J. D. (2005). *Logic, meaning, and conversation: Semantical underdeterminacy, implicature, and their interface*. New York: Oxford University Press.
- Austin, J. L. (1966). "Three ways of spilling ink". In: *The Philosophical Review* 75.4, pp. 427–440.
- Austin, J. L. (1962). *How to do Things with Words. The William James lectures delivered at Harvard University in 1955*. Ed. by J. O. Urmson. Clarendon Press.
- Bangerter, A. and H. H. Clark (2003). "Navigating joint projects with dialogue". In: *Cognitive Science* 27.2, pp. 195–225.
- Bangerter, A. (2004). "Using pointing and describing to achieve joint focus of attention in dialogue". In: *Psychological Science* 15.6, pp. 415–419.
- Bangerter, A., H. H. Clark, and A. R. Katz (2004). "Navigating joint projects in telephone conversations". In: *Discourse Processes* 37.1, pp. 1–23.
- Bara, B. G. (2010). *Cognitive pragmatics: The mental processes of communication*. MIT Press.
- (2011). "Cognitive pragmatics: The mental processes of communication". In: *Intercultural Pragmatics* 8.3, pp. 443–485.

- Barr, D. J. et al. (2013). “Random effects structure for confirmatory hypothesis testing: Keep it maximal”. In: *Journal of memory and language* 68.3, pp. 255–278.
- Bates, D. et al. (2015). “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67.1, pp. 1–48.
- Bavelas, J. and S. Healing (2013). “Reconciling the effects of mutual visibility on gesturing: A review”. In: *Gesture* 13.1, pp. 63–92.
- Bavelas, J. et al. (2011). “Dyadic evidence for grounding with abstract deictic gestures”. In: *Integrating Gestures: The Interdisciplinary Nature of Gesture*. Ed. by G. Stam and M. Ishino. Vol. 4. Gesture Studies. Amsterdam: John Benjamins Publishing Company, pp. 49–60.
- Bavelas, J. et al. (2008). “Gesturing on the telephone: Independent effects of dialogue and visibility”. In: *Journal of Memory and Language* 58.2, pp. 495–520.
- Bavelas, J. et al. (2012). “Beyond Back-channels: A Three-step Model of Grounding in Face-to-face Dialogue”. In: *Proceedings of Interdisciplinary Workshop on Feedback Behaviors in Dialog*, pp. 5–6.
- Beattie, G. and H. Shovelton (2002). “An experimental investigation of some properties of individual iconic gestures that mediate their communicative power”. In: *British Journal of Psychology* 93, pp. 179–192.
- Beattie, G. and H. Shovelton (1999a). “Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation”. In: *Semiotica* 123.1-2, pp. 1–30.
- (1999b). “Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech”. In: *Journal of Language and Social Psychology* 18.4, pp. 438–462.
- (2001). “An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach”. In: *Gesture* 1.2, pp. 129–149.
- (2006). “When size really matters: How a single semantic feature is represented in the speech and gesture modalities”. In: *Gesture* 6.1, pp. 63–84.
- Behne, T., M. Carpenter, and M. Tomasello (2005). “One-year-olds comprehend the communicative intentions behind gestures in a hiding game”. In: *Developmental science* 8.6, pp. 492–499.
- Benz, A., G. Jäger, and R. van Rooij (2006). *Game Theory and Pragmatics*. Palgrave Studies in Pragmatics, Languages and Cognition. Palgrave Macmillan. ISBN: 9781403945723.
- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Vol. 99. Cambridge University Press.
- Blutner, R., H. de Hoop, and P. Hendriks (2006). *Optimal communication*. CSLI lecture notes. CSLI Publications. ISBN: 9781575865133.
- Blutner, R. and H. Zeevat (2004). *Optimality Theory and Pragmatics*. Palgrave Studies in Pragmatics, Language and Cognition. Palgrave Macmillan Limited. ISBN: 9781403901293.
- Brennan, S. E. and H. H. Clark (1996). “Conceptual pacts and lexical choice in conversation”. In: *Journal of Experimental Psychology-Learning Memory and Cognition* 22.6, pp. 1482–1493.
- Brown, G. (1995). *Speakers, listeners and communication: Explorations in discourse analysis*. Cambridge: Cambridge University Press.
- Brown-Schmidt, S. (2009). “Partner-specific interpretation of maintained referential precedents during interactive dialog”. In: *Journal of Memory and Language* 61.2, pp. 171–190.
- Calbris, G. (1990). *The semiotics of French gestures*. Indiana University Press Bloomington.
- Carassa, A. and M. Colombetti (2009). “Joint meaning”. In: *Journal of Pragmatics* 41.9, pp. 1837–1854.

- Carletta, J., A. Isard, and J. Kowtko (1996). *HCRC dialogue structure coding manual*. HCRC Publications, University of Edinburgh.
- Carletta, J. et al. (1997). “The reliability of a dialogue structure coding scheme”. In: *Computational Linguistics* 23.1, pp. 13–31.
- Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Wiley New York.
- (2008). “Linguistic communication and the semantics/pragmatics distinction”. In: *Synthese* 165.3, pp. 321–345.
- Chambers, C. G., M. K. Tanenhaus, and J. S. Magnuson (2004). “Actions and affordances in syntactic ambiguity resolution.” In: *Journal of experimental psychology: Learning, memory, and cognition* 30.3, p. 687.
- Chomsky, N. (2010). “Some simple evo devo theses: How true might they be for language”. In: *The evolution of human language: biolinguistic perspective*. Ed. by R. K. Larson, V. M. Déprez, and H. Yamakido. Cambridge: Cambridge University Press, pp. 45–62.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- (2003). “Pointing and placing.” In: *Pointing: Where Language, Culture, and Cognition Meet*. Ed. by S. Kita. Lawrence Erlbaum Associates Publishers, pp. 243–268.
- Clark, H. H. and S. E. Brennan (1991). “Grounding in communication”. In: *Perspectives on socially shared cognition*. Ed. by L. B. Resnick, J. M. Levine, and S. D. Teasley. Vol. 13. Washington DC: American Psychological Association, pp. 127–149.
- Clark, H. H. and M. A. Krych (2004). “Speaking while monitoring addressees for understanding”. In: *Journal of Memory and Language* 50.1, pp. 62–81.
- Clark, H. H. and E. F. Schaefer (1989). “Contributing to discourse”. In: *Cognitive science* 13.2, pp. 259–294.
- Clark, H. H. and D. Wilkes-Gibbs (1986). “Referring as a collaborative process”. In: *Cognition* 22.1, pp. 1–39.
- Cohen, D., G. Beattie, and H. Shovelton (2011). “Tracking the distribution of individual semantic features in gesture across spoken discourse: New perspectives in multimodal interaction”. In: *Semiotica* 185.1-4, pp. 147–188.
- Cohen, R. L. and N. Otterbein (1992). “The mnemonic effect of speech gestures: Pantomimic and non-pantomimic gestures compared”. In: *European Journal of Cognitive Psychology* 4.2, pp. 113–139.
- Cook, S. W., Z. Mitchell, and S. Goldin-Meadow (2008). “Gesturing makes learning last”. In: *Cognition* 106.2, pp. 1047–1058.
- Cooper, R. M. (1974). “The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing”. In: *Cognitive Psychology* 6.1, pp. 84–107.
- Crystal, D. and D. Davy (1969). *Investigating English Style*. ERIC.
- Dahan, D. and M. K. Tanenhaus (2005). “Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition”. In: *Psychonomic Bulletin & Review* 12.3, pp. 453–459.
- De Ruiter, J. P. (2007). “Postcards from the mind: The relationship between speech, imagistic gesture, and thought”. In: *Gesture* 7.1, pp. 21–38.
- De Ruiter, J. P. (2000). “The production of gesture and speech”. In: *Language and gesture*. Ed. by D. McNeill. Cambridge University Press, pp. 284–311.
- De Ruiter, J. P., A. Bangerter, and P. Dings (2012a). “The Interplay Between Gesture and Speech in the Production of Referring Expressions: Investigating the Tradeoff Hypothesis”. In: *Topics in Cognitive Science* 4.2, pp. 232–248.

- De Ruiter, J. P., A. Bangerter, and P. Dings (2012b). “The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis”. In: *Topics in cognitive science* 4.2, pp. 232–248.
- Dennett, D. C. (1993). *Consciousness explained*. Penguin UK.
- Driskell, J. E. and P. H. Radtke (2003). “The effect of gesture on speech production and comprehension”. In: *Human Factors* 45.3, pp. 445–454.
- Eberhard, K. M. et al. (1995). “Eye movements as a window into real-time spoken language comprehension in natural contexts”. In: *Journal of psycholinguistic research* 24.6, pp. 409–436.
- Ekman, P. and W. V. Friesen (1969). “The repertoire of nonverbal behavior: Categories, origins, usage, and coding”. In: *Semiotica* 1.1, pp. 49–98.
- Emmorey, K. and S. Casey (2002). “Gesture, Thought, and Spatial Language”. In: *Spatial Language: Cognitive and Computational Perspectives*. Ed. by K. R. Coventry and P. Olivier. Kluwer Academic Publishers, pp. 87–102.
- Enfield, N. J. (2001). “Lip-pointing: A discussion of form and function with reference to data from Laos”. In: *Gesture* 1.2, pp. 185–211.
- (2004). “On linear segmentation and combinatorics in co-speech gesture: A symmetry-dominance construction in Lao fish trap descriptions”. In: *Semiotica* 149, pp. 57–124.
- (2009a). “Relationship thinking and human pragmatics”. In: *Journal of Pragmatics* 41.1, pp. 60–78.
- (2009b). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.
- Enfield, N. (2013). “A ‘Composite Utterances’ approach to meaning.” In: *Handbook of Body – Language – Communication*. Ed. by C. Müller et al. Vol. 1. Berlin: Mouton de Gruyter.
- Engelhardt, P. E., K. G. Bailey, and F. Ferreira (2006). “Do speakers and listeners observe the Gricean Maxim of Quantity?” In: *Journal of Memory and Language* 54.4, pp. 554–573.
- Feyereisen, P. (2006). “Further investigation on the mnemonic effect of gestures: Their meaning matters”. In: *European Journal of Cognitive Psychology* 18.2, pp. 185–205.
- Frith, C. and U. Frith (2005). “Theory of mind”. In: *Current Biology* 15.17, R644–R645.
- Galati, A. and A. b. c. Samuel (2011). “The role of speech-gesture congruency and delay in remembering action events”. In: *Language and Cognitive Processes* 26.3. cited By 1, pp. 406–436. DOI: 10.1080/01690965.2010.494846.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Prentice Hall.
- Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. Academic Press New York.
- Gergely, G., H. Bekkering, and I. Király (2002). “Developmental psychology: rational imitation in preverbal infants”. In: *Nature* 415.6873, pp. 755–755.
- Gerwing, J. and M. Allison (2009a). “The relationship between verbal and gestural contributions in conversation A comparison of three methods”. In: *Gesture* 9.3, pp. 312–336.
- Gerwing, J. and M. Allison (2009b). “The relationship between verbal and gestural contributions in conversation: A comparison of three methods”. In: *Gesture* 9.3, pp. 312–336.
- Gerwing, J. and J. Bavelas (2004). “Linguistic influences on gesture’s form”. In: *Gesture* 4.2, pp. 157–195.
- Gigerenzer, G., P. M. Todd, and ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.



- Goffman, E. (1964). “The neglected situation”. In: *American anthropologist* 66.6(2), pp. 133–136.
- (1981). “Footing”. In: *Forms of talk*. University of Pennsylvania Press, Philadelphia, pp. 124–159.
- Goldin-Meadow, S. et al. (2001). “Explaining math: Gesturing lightens the load”. In: *Psychological Science* 12.6, pp. 516–522.
- Goldstein, D. G. and G. Gigerenzer (2002). “Models of ecological rationality: the recognition heuristic.” In: *Psychological review* 109.1, p. 75.
- Graham, J. A. and S. Heywood (1975). “The effects of elimination of hand gestures and of verbal codability on speech performance”. In: *European Journal of Social Psychology* 5.2, pp. 189–195.
- Grassmann, S. and M. Tomasello (2010). “Young children follow pointing over words in interpreting acts of reference”. In: *Developmental science* 13.1, pp. 252–263.
- Grice, H. P. (1957). “Meaning”. In: *The philosophical review* 66.3, pp. 377–388.
- (1975). “Logic and conversation”. In: *Syntax and Semantics 3: Speech Acts*. Ed. by P. Cole and J. Morgan. New York: Academic Press, pp. 41–58.
- Grice, H. (1989). *Study in the way of words*. Cambridge, MA: Harvard University Press.
- Hallett, P. (1986). “Eye movements((and human visual perception))”. In: *Handbook of perception and human performance*. 1, pp. 10.11–10.112.
- Hanna, J. E. and S. E. Brennan (2007). “Speakers’ eye gaze disambiguates referring expressions early during face-to-face conversation”. In: *Journal of Memory and Language* 57.4, pp. 596–615.
- Haviland, J. B. (1993). “Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures”. In: *Journal of linguistic anthropology* 3.1, pp. 3–45.
- Heritage, J. (2012). “Epistemics in Action: Action Formation and Territories of Knowledge”. In: *Research on Language and Social Interaction* 45.1, pp. 1–29.
- Hewes, G. W. (1981). “Pointing and language”. In: *Advances in Psychology* 7, pp. 263–269.
- Hinojosa, J. A., M. Martín-Loeches, and F. J. Rubia (2001). “Event-related potentials and semantics: An overview and an integrative proposal”. In: *Brain and Language* 78.1, pp. 128–139.
- Hoetjes, M., E. Kraemer, and M. Swerts (2014). “Does our speech change when we cannot gesture?” In: *Speech Communication* 57, pp. 257–267.
- Holle, H. and T. C. Gunter (2007). “The role of iconic gestures in speech disambiguation: ERP evidence”. In: *Journal of cognitive neuroscience* 19.7, pp. 1175–1192.
- Holler, J. b., K. Turner, and T. Varciana (2013). “It’s on the tip of my fingers: Co-speech gestures during lexical retrieval in different social contexts”. In: *Language and Cognitive Processes* 28.10. cited By 1, pp. 1509–1518. DOI: 10 . 1080 / 01690965 . 2012 . 698289.
- Holler, J. and G. Beattie (2002). “A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk”. In: *Semiotica* 142.1/4, pp. 31–70.
- (2003a). “How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process?” In: *Semiotica* 146.1/4, pp. 81–116.
- (2003b). “Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener?” In: *Gesture* 3.2, pp. 127–154.
- Holler, J. and R. Stevens (2007). “The effect of common ground on how speakers use gesture and speech to represent size information”. In: *Journal of Language and Social Psychology* 26.1, pp. 4–27.
- Holler, J. and K. Wilkin (2009). “Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task”. In: *Language and cognitive processes* 24.2, pp. 267–289.

- Hostetter, A. B. and M. W. Alibali (2005). *On the Tip of the Mind: Gesture as a Key to Conceptualization*. Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society, pp. 589–594.
- Hostetter, A. B. (2011). “When do gestures communicate? A meta-analysis.” In: *Psychological bulletin* 137.2, pp. 297–315.
- Hostetter, A. B. and M. W. Alibali (2007). “Raise your hand if you’re spatial: Relations between verbal and spatial skills and gesture production”. In: *Gesture* 7.1, pp. 73–95.
- Hostetter, A. B., M. W. Alibali, and S. Kita (2007). “Does sitting on your hands make you bite your tongue? The effects of gesture prohibition on speech during motor descriptions”. In: *Proceedings of the 29th annual meeting of the Cognitive Science Society*. Ed. by D. S. McNamara and J. G. Trafton. Mahwah, NJ: Erlbaum, pp. 1097–1102.
- Huettig, F., J. Rommers, and A. S. Meyer (2011). “Using the visual world paradigm to study language processing: A review and critical evaluation”. In: *Acta Psychologica* 137.2, pp. 151–171.
- Huettig, F. and G. T. Altmann (2007). “Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness”. In: *Visual Cognition* 15.8, pp. 985–1018.
- Huettig, F. and J. M. McQueen (2007). “The tug of war between phonological, semantic and shape information in language-mediated visual search”. In: *Journal of Memory and Language* 57.4, pp. 460–482.
- Isard, A. and J. Carletta (1995). *Transaction and action coding in the Map Task Corpus*. CiteSeer.
- Iverson, J. M. and S. Goldin-Meadow (2001). “The resilience of gesture in talk: gesture in blind speakers and listeners”. In: *Developmental Science* 4.4, pp. 416–422.
- Iverson, J. M. et al. (2000). “The relation between gesture and speech in congenitally blind and sighted language-learners”. In: *Journal of nonverbal behavior* 24.2, pp. 105–130.
- Jacobs, N. and A. Garnham (2007). “The role of conversational hand gestures in a narrative task”. In: *Journal of Memory and Language* 56.2, pp. 291–303.
- Johnston, T. and A. Schembri (2007). *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge University Press.
- Kelly, S. D., C. Kravitz, and M. Hopkins (2004). “Neural correlates of bimodal speech and gesture comprehension”. In: *Brain and Language* 89.1, pp. 253–260.
- Kelly, S. D. et al. (1999). “Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory”. In: *Journal of Memory and Language* 40.4, pp. 577–592.
- Kelly, S. b. et al. (2015). “The processing of speech, gesture, and action during language comprehension”. In: *Psychonomic Bulletin and Review* 22.2, pp. 517–523.
- Kelly, S. D., A. Özyürek, and E. Maris (2010). “Two Sides of the Same Coin Speech and Gesture Mutually Interact to Enhance Comprehension”. In: *Psychological Science* 21.2, pp. 260–267.
- Kendon, A. (1991). “Some considerations for a theory of language origins”. In: *Man*, pp. 199–221.
- (1994). “Do gestures communicate? A review”. In: *Research on language and social interaction* 27.3, pp. 175–200.
- (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kissine, M. (2013). *From utterances to speech acts*. Cambridge University Press.
- Kita, S. (2009). “Cross-cultural variation of speech-accompanying gesture: A review”. In: *Language and Cognitive Processes* 24.2, pp. 145–167.
- Kita, S. and T. S. Davies (2009). “Competing conceptual representations trigger co-speech representational gestures”. In: *Language and Cognitive Processes* 24.5, pp. 761–775.

- Kita, S. (1993). “Language and thought interface: A study of spontaneous gestures and Japanese mimetics”. PhD thesis. University of Chicago Chicago, IL.
- (2000). “How representational gestures help speaking”. In: *Language and gesture*. Ed. by D. McNeill. Cambridge University Press, pp. 162–185.
- Kita, S. and A. Özyürek (2003). “What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking”. In: *Journal of Memory and language* 48.1, pp. 16–32.
- Kita, S. et al. (2007). “Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production”. In: *Language and Cognitive Processes* 22.8, pp. 1212–1236.
- Kockelman, P. (2005). “The semiotic stance”. In: *Semiotica* 2005.157, pp. 233–304.
- (2012). “Meaning, motivation, and mind: Some conditions and consequences for the flexibility and intersubjectivity of cognitive processes”. In: *New Ideas in Psychology* 30.1, pp. 65–85.
- Kok, K. I. and A. Cienki (2016). “Cognitive Grammar and gesture: Points of convergence, advances and challenges”. In: *Cognitive Linguistics* 27.1, pp. 67–100.
- Kowtko, J. C., S. D. Isard, and G. M. Doherty (1993). “Conversational games within dialogue”. In: *University of Edinburgh*. Citeseer.
- Krauss, R. M., P. Morrel-Samuels, and C. Colasante (1991). “Do Conversational Hand Gestures Communicate?” In: *Journal of Personality and Social Psychology* 61.5, pp. 743–754.
- Krauss, R. M., Y. Chen, and R. F. Gottesmann (2000). “Lexical gestures and lexical access: a process model”. In: *Language and gesture*. Ed. by D. McNeill. Vol. 2. Cambridge University Press, pp. 261–283.
- Krauss, R. M. et al. (1995). “The communicative value of conversational hand gesture”. In: *Journal of Experimental Social Psychology* 31.6, pp. 533–552.
- Krauss, R. M. et al. (1999). “The role of speech-related arm/hand gestures in word retrieval”. In: *Gesture, Speech, and Sign*. Ed. by M. L. S. and C. R. Oxford University Press, pp. 93–116.
- Kronmuller, E. and D. J. Barr (2007). “Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis”. In: *Journal of Memory and Language* 56.3, pp. 436–455.
- Kutas, M. and S. A. Hillyard (1980). “Event-related brain potentials to semantically inappropriate and surprisingly large words”. In: *Biological Psychology* 11.2, pp. 99–116.
- Langacker, R. W. (2013). *Essentials of cognitive grammar*. Oxford University Press.
- Levelt, W. J. (1993). *Speaking: From intention to articulation*. Vol. 1. MIT press.
- Levinson, S. C. (1979). “Activity types and language”. In: *Linguistics* 17.5-6, pp. 365–400.
- (1983). *Pragmatics*. Cambridge University Press.
- (2000). *Presumptive meanings: The theory of generalized conversational implicature*. The MIT Press.
- (2003). *Space in language and cognition: Explorations in cognitive diversity*. Vol. 5. Cambridge University Press.
- Lewis, D. (1969). *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.
- Loucks, J. and E. Pederson (2010). “Linguistic and non-linguistic categorization of complex motion events”. In: *Event Representation in Language and Cognition*. Ed. by J. Bohnemeyer and E. Pederson. Cambridge University Press.
- Lücking, A. et al. (2013). “Data-based analysis of speech and gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications”. In: *Journal on Multimodal User Interfaces* 7.1-2, pp. 5–18.

- Matin, E., K. Shao, and K. R. Boff (1993). “Saccadic overhead: Information-processing time with and without saccades”. In: *Perception & Psychophysics* 53.4, pp. 372–380.
- McNeill, D. (2009). “Gesture and Communication”. In: *Concise Encyclopedia of Pragmatics*. Ed. by J. L. Mey. 2nd. Oxford: Elsevier, pp. 299–307.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- (2005). *Gesture and thought*. London: University of Chicago Press.
- (2012). *How language began: Gesture and speech in human evolution*. Cambridge University Press.
- (2015). *Why We Gesture: The Surprising Role of Hand Movements in Communication*. Cambridge University Press.
- McNeill, D. and S. D. Duncan (2000). “Growth points in speaking-for-thinking”. In: *Language and gesture*. Ed. by D. McNeill. Cambridge: Cambridge University Press, pp. 141–161.
- Mehler, A., A. Lücking, and P. Weiss (2010). “A Network Model of Interpersonal Alignment in Dialog”. In: *Entropy* 12.6, pp. 1440–1483.
- Melinger, A. and S. Kita (2007). “Conceptualisation load triggers gesture production”. In: *Language and Cognitive Processes* 22.4, pp. 473–500.
- Melinger, A. and W. J. Levelt (2004). “Gesture and the communicative intention of the speaker”. In: *Gesture* 4.2, pp. 119–141.
- Metzing, C. and S. E. Brennan (2003). “When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions”. In: *Journal of Memory and Language* 49.2, pp. 201–213.
- Mol, L. et al. (2011). “Seeing and Being Seen: The Effects on Gesture Production”. In: *Journal of Computer-Mediated Communication* 17.1, pp. 77–100.
- Morsella, E. and R. M. Krauss (2004). “The role of gestures in spatial working memory and speech”. In: *American Journal of Psychology* 117.3, pp. 411–424.
- Mushin, I. et al. (2003). “Discourse structure, grounding, and prosody in task-oriented dialogue”. In: *Discourse Processes* 35.1, pp. 1–31.
- Neider, M. B. et al. (2010). “Coordinating spatial referencing using shared gaze”. In: *Psychonomic Bulletin & Review* 17.5, pp. 718–724.
- Novack, M. and S. Goldin-Meadow (2015). “Learning from gesture: how our hands change our minds”. In: *Educational Psychology Review* 27.3, pp. 405–412.
- Özyürek, A. (2014). “Hearing and seeing meaning in speech and gesture: insights from brain and behaviour”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1651, p. 20130296.
- Özyürek, A. b. et al. (2005). “How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities”. In: *Gesture* 5.1-2, pp. 219–240.
- Özyürek, A. (2002). “Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures”. In: *Journal of Memory and Language* 46.4, pp. 688–704.
- Özyürek, A. et al. (2007). “On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials”. In: *Journal of Cognitive Neuroscience* 19.4, pp. 605–616.
- Parikh, P. (2010). *Language and equilibrium*. MIT Press.
- Peirce, C. S. (1998). *The essential Peirce: Selected philosophical writings, 1893-1913*. Vol. 1. Indiana University Press.
- Peirce, C. S. S. (1955). *Philosophical Writings of Peirce*. Ed. by J. Buchler. Dover Publications.

- Rauscher, F. H., R. M. Krauss, and Y. S. Chen (1996). “Gesture, speech, and lexical access: The role of lexical movements in speech production”. In: *Psychological Science* 7.4, pp. 226–231.
- Recanati, F. (2004). *Literal meaning*. Cambridge University Press.
- (2010). *Truth-conditional pragmatics*. Clarendon Press Oxford.
- Richardson, D. C., R. Dale, and J. M. Tomlinson (2009). “Conversation, Gaze Coordination, and Beliefs About Visual Context”. In: *Cognitive Science* 33.8, pp. 1468–1482.
- Rizzolatti, G. and M. A. Arbib (1998). “Language within our grasp”. In: *Trends in neurosciences* 21.5, pp. 188–194.
- Sacks, H. and E. A. Schegloff (2002). “Home position”. In: *Gesture* 2.2, pp. 133–146.
- Salverda, A. P. and M. K. Tanenhaus (2010). “Tracking the time course of orthographic information in spoken-word recognition.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36.5, p. 1108.
- Schegloff, E. A. (1968). “Sequencing in Conversational Openings<sup>1</sup>”. In: *American anthropologist* 70.6, pp. 1075–1095.
- (1984). “On some gestures’ relation to talk”. In: *Structures of Social Action: Studies in Conversation Analysis*. Ed. by J. Atkinson and E. J. Heritage. Cambridge University Press Cambridge, England, pp. 266–296.
- Schegloff, E. A., G. Jefferson, and H. Sacks (1977). “The preference for self-correction in the organization of repair in conversation”. In: *Language*, pp. 361–382.
- Schegloff, E. A. and H. Sacks (1973). “Opening up closings”. In: *Semiotica* 8.4, pp. 289–327.
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge University Press.
- Schober, M. F. and H. H. Clark (1989). “Understanding by addressees and overhearers”. In: *Cognitive Psychology* 21.2, pp. 211–232.
- Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.
- (2010). *Making the social world: The structure of human civilization*. Oxford University Press.
- Short, T. L. (2007). *Peirce’s theory of signs*. Cambridge University Press.
- Silverman, L. B. et al. (2010). “Speech-and-gesture integration in high functioning autism”. In: *Cognition* 115.3, pp. 380–393.
- Slobin, D. I. (1987). “Thinking for speaking”. In: *Annual Meeting of the Berkeley Linguistics Society*. Vol. 13, pp. 435–445.
- So, W. C., S. Kita, and S. Goldin-Meadow (2009). “Using the Hands to Identify Who Does What to Whom: Gesture and Speech Go Hand-in-Hand”. In: *Cognitive science* 33.1, pp. 115–125.
- So, W. et al. (2013). “Iconic gestures prime words: comparison of priming effects when gestures are presented alone and when they are accompanying speech.” In: *Frontiers in psychology* 4.
- Sperber, D. (1995). “How do we communicate”. In: *How things are: A science toolkit for the mind*, pp. 191–199.
- Sperber, D. and D. Wilson (2012). “The mapping between the mental and the public lexicon”. In: *Meaning and Relevance*. Ed. by D. Wilson and D. Sperber. Cambridge: Cambridge University Press, pp. 31–46.
- Sperber, D. and D. Wilson (1995). *Relevance: Communication and Cognition*. Second. Blackwell Publishing, Oxford.
- Sperber, D. and D. Wilson (1986). *Relevance: Communication and cognition*. Harvard University Press Cambridge, MA.
- (1987). “Precis of relevance: communication and cognition”. In: *Behavioral and brain sciences* 10.4, pp. 697–754.

- Sperber, D. and D. Wilson (2002). “Pragmatics, Modularity and Mind-reading”. In: *Mind & Language* 17.1-2, pp. 3–23.
- Stokoe, W. C. (2005[1960]). “Sign language structure: An outline of the visual communication systems of the American deaf”. In: *Journal of deaf studies and deaf education* 10.1, pp. 3–37.
- Streeck, J. (2008). “Depicting by gesture”. In: *Gesture* 8.3, pp. 285–301.
- (2009). *Gesturecraft: The manufacture of meaning*. Vol. 2. John Benjamins Publishing.
- Sutton-Spence, R. and B. Woll (1999). *The linguistics of British Sign Language: an introduction*. Cambridge University Press.
- Talmy, L. (2000a). *Toward a cognitive semantics, vol. I: Concept structuring systems*. Cambridge, MA: MIT Press.
- (2000b). *Toward a cognitive semantics, vol. II: Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press.
- Tanenhaus, M. K. et al. (1995). “Integration of visual and linguistic information in spoken language comprehension”. In: *Science* 268.5217, pp. 1632–1634.
- Tomasello, M. (2010). *Origins of human communication*. MIT press.
- Tomasello, M., M. Carpenter, and U. Liszkowski (2007). “A new look at infant pointing”. In: *Child development* 78.3, pp. 705–722.
- Tomasello, M. et al. (2005). “Understanding and sharing intentions: The origins of cultural cognition”. In: *Behavioral and brain sciences* 28.5, pp. 675–690.
- Tversky, B. et al. (2009). “Explanations in gesture, diagram, and word”. In: *Spatial Language and Dialogue*. Ed. by K. R. Coventry, T. Tenbrink, and J. A. Bateman. Oxford University Press Oxford, pp. 119–131.
- Tyler, A. and V. Evans (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition*. Cambridge University Press.
- Wagner, S. M., H. Nusbaum, and S. Goldin-Meadow (2004). “Probing the mental representation of gesture: Is handwaving spatial?” In: *Journal of Memory and Language* 50.4, pp. 395–407.
- Watson, J. and J. Wilson (Submitted). “Gesture in Modern South Arabian Languages: variation in multimodal constructions during task-based interaction”. In: *Brill’s Journal of Afroasiatic Languages and Linguistics*.
- Wesp, R. et al. (2001). “Gestures maintain spatial imagery”. In: *American Journal of Psychology* 114.4, pp. 591–600.
- Wharton, T. (2003). “Natural pragmatics and natural codes”. In: *Mind & language* 18.5, pp. 447–477.
- (2009). *Pragmatics and non-verbal communication*. Cambridge: Cambridge University Press.
- Wilson, D. (2003). “Relevance and lexical pragmatics”. In: *Italian Journal of Linguistics* 15, pp. 273–292.
- Wilson, J. J. (2011). “Contextual Substitutes: Exploring the role of gestural sequences during interaction”. MA thesis. University College London.
- Winter, B. (2013). *A very basic tutorial for performing linear mixed effects analyses*. URL: [http://www.bodowinter.com/tutorial/bw\\_LME\\_tutorial.pdf](http://www.bodowinter.com/tutorial/bw_LME_tutorial.pdf) (visited on 2013).
- (2015). “The other N: The role of repetitions and items in the design of phonetic experiments”. In: *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.
- Wittgenstein, L. (1958). *The Blue and Brown Books*. Basil Blackwell, Oxford.

- Wu, Y. b. and S. c. Coulson (2014). “Co-speech iconic gestures and visuo-spatial working memory”. In: *Acta Psychologica* 153, pp. 39–50. doi: 10 . 1016 / j . actpsy . 2014 . 09 . 002.
- Wu, Y. C. and S. Coulson (2005). “Meaningful gestures: Electrophysiological indices of iconic gesture comprehension”. In: *Psychophysiology* 42.6, pp. 654–667.
- Yap, D. F. et al. (2011). “Iconic Gestures Prime Words”. In: *Cognitive Science* 35.1, pp. 171–183.
- Yasui, E. (2013). “Collaborative idea construction: Repetition of gestures and talk in joint brainstorming”. In: *Journal of Pragmatics*, pp. 157–172.
- Zima, E. (2014). “English multimodal motion constructions. A construction grammar perspective”. In: *Linguistic Society of Belgium* 8.





# Appendix

## Map Task Instructions

As part of this study you are required to complete two collaborative tasks.

### Task One

For the first task, both of you will have a workspace containing a collection of two dimensional shapes, and your job is to cooperate in order to make sure that they are in the same order.

YOU CANNOT SHOW EACH OTHER YOUR WORKSPACES

Once you have completed the task, please contact the experimenter.

To begin the task please uncover your work space.

Do you have any questions?

### Task Two

For the second task, this time your workspaces will contain two dimensional 'treasure' maps. These maps are NOT be the same. On your maps, one of you will have ten landmarks, a start & end point, and a route, while the other will only have 10 landmarks, and a start point. Your task is to cooperate in order ensure that both maps have routes and end points.

Once again, YOU CANNOT SHOW EACH OTHER YOUR WORKSPACES.

Once you have completed the task, please contact the experimenter.

To begin the task please pull the tab at the top of the page in front of you.

Do you have any questions?

## Eye tracking Consent Form

As part of my research you have agreed to participate in an experiment task. During the task, you're decisions and eye movements will be recorded. These recordings may be used in a variety of ways, each of which require your consent. Please indicate your consent below. This is completely up to you. I will only use the records in ways that you agree to. In any use of these records, names will not be identified.

In each of the following, please sign your initials to show your agreement:

1. The records can be studied by the research team for use in the research project.
2. The records can be shown to subjects in other experiments.
3. The records can be used for scientific publications.
4. The records can be used by other researchers.
5. The records can be shown at meetings of scientists interested in the study of Language.
6. The records can be shown in classrooms to students.
7. The records can be shown in public presentations to non-scientific groups.
8. The records can be used on television and radio.

I have read the above description and give my consent for the use of the records as indicated above.

Date

Signature

Name

Researcher contact details:

Jack J. Wilson  
PhD Student  
Linguistics and Phonetics  
School of Languages, Cultures and Societies  
University of Leeds  
Leeds LS2 9JTk

## Map Task Consent Form

As part of my research you have agreed to participate in a conversational task. During the task you will be both video and audio recorded. These recordings may be used in a variety of ways, each of which require your consent. Please indicate your consent below. This is completely up to you. I will only use the records in ways that you agree to. In any use of these records, names will not be identified.

In each of the following, please sign your initials to show your agreement:

1. The records can be studied by the research team for use in the research project.
2. The records can be shown to subjects in other experiments.
3. The records can be used for scientific publications.
4. The written transcript can be kept in an archive for other researchers.
5. The records can be used by other researchers.
6. The records can be shown at meetings of scientists interested in the study of Language.
7. The records can be shown in classrooms to students.
8. The records can be shown in public presentations to non-scientific groups.
9. The records can be used on television and radio.

I have read the above description and give my consent for the use of the records as indicated above.

Date

Signature

Name

Researcher contact details:

Jack J. Wilson  
Provisional PhD Student  
Linguistics and Phonetics  
School of Modern Languages & Cultures  
University of Leeds  
Leeds LS2 9JT

## Map Task R script

```

data <- read.csv("~/GoogleDrive/DB/0_phdWork/PhD_
  ↪ fromComputer/Data/Maptask/Stats/MT_final.csv", header
  ↪ = T)
summary(data)

lmerControl(optimizer="bobyqa", optCtrl = list(maxfun =
  ↪ 10000))
#####
# INTRODUCTION. PARTICIPANTS AND MOVE TYPE
#####
# The effect of participants on gesture
#####

null <- glmer(Gesture ~ 1 + (1|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Gesture <- glmer(Gesture ~ Participant + (1|Game_Coding_
  ↪ Label), data = data, family = binomial)

anova(null, Gesture)
Data: data
Models:
  null: Gesture ~ 1 + (1 | Game_Coding_Label)
Gesture: Gesture ~ Participant + (1 | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
null    2 2670.6 2681.8 -1333.3    2666.6
Gesture 17 2271.7 2367.1 -1118.8    2237.7 428.88    15 <
  ↪ 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Gesture ~ Participant + (1 | Game_Coding_Label)
Data: data

AIC      BIC    logLik deviance df.resid
2271.7    2367.1  -1118.8    2237.7      2004

Scaled residuals:
  Min      1Q  Median      3Q      Max
-4.9469 -0.5728 -0.4416  0.8304  3.3001

Random effects:

```

```

Groups              Name              Variance Std.Dev.
Game_Coding_Label (Intercept) 0.004517 0.06721
Number of obs: 2021, groups: Game_Coding_Label, 16

Fixed effects:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.12918    0.12099  -9.333 < 2e-16 ***
ParticipantP02 -0.87426    0.27586  -3.169 0.00153 **
ParticipantP03  2.14661    0.22909   9.370 < 2e-16 ***
ParticipantP04  0.96081    0.36192   2.655 0.00794 **
ParticipantP05  4.31199    0.73136   5.896 3.73e-09 ***
ParticipantP06 -0.46103    0.64942  -0.710 0.47776
ParticipantP07  1.36318    0.18511   7.364 1.78e-13 ***
ParticipantP08  1.42875    0.24281   5.884 4.00e-09 ***
ParticipantP09  0.41967    0.42808   0.980 0.32691
ParticipantP10  0.88085    0.51952   1.696 0.08998 .
ParticipantP11 -0.09739    0.24141  -0.403 0.68664
ParticipantP12  1.37421    0.33480   4.105 4.05e-05 ***
ParticipantP13 -1.23471    0.41264  -2.992 0.00277 **
ParticipantP14  0.63177    0.53420   1.183 0.23695
ParticipantP15  1.50407    0.16090   9.348 < 2e-16 ***
ParticipantP16 -0.48159    0.21483  -2.242 0.02498 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                ↪ ' ' 1

# Participant has a significant effect on the production of
  ↪ gesture.

#####
# The effect of Move type on gesture
#####
gc.null <- glmer(Gesture ~ 1 + (1|Participant), data = data
  ↪ , family = binomial)
gc.Gesture <- glmer(Gesture ~ Game_Coding_Label + (1|
  ↪ Participant), data = data, family = binomial)
anova(gc.null, gc.Gesture)
Data: data
Models:
  gc.null: Gesture ~ 1 + (1 | Participant)
  gc.Gesture: Gesture ~ Game_Coding_Label + (1 | Participant)
Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
gc.null      2 2307.3 2318.6 -1151.7 2303.3
gc.Gesture 14 2309.8 2388.3 -1140.9 2281.8 21.566 12
  ↪ 0.04268 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                ↪ ' ' 1

summary(gc.Gesture)

```

```

Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Gesture ~ Game_Coding_Label + (1 | Participant)
Data: data

AIC      BIC    logLik deviance df.resid
2309.8   2388.3 -1140.9  2281.8    2007

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.5160 -0.5861 -0.4288  0.8009  3.4346

Random effects:
  Groups      Name      Variance Std.Dev.
Participant (Intercept) 1.462    1.209
Number of obs: 2021, groups: Participant, 16

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)                -0.538439  0.423796 -1.270
  ↪ 0.2039
Game_Coding_LabelAlign      0.448058  0.339788  1.319
  ↪ 0.1873
Game_Coding_LabelCheck     -0.096369  0.319328 -0.302
  ↪ 0.7628
Game_Coding_LabelClarify    0.009167  0.329963  0.028
  ↪ 0.9778
Game_Coding_LabelExp       -0.086730  0.343949 -0.252
  ↪ 0.8009
Game_Coding_LabelInst       0.252832  0.320877  0.788
  ↪ 0.4307
Game_Coding_LabelQ-Wh       0.001153  0.584261  0.002
  ↪ 0.9984
Game_Coding_LabelQ-Y/N      0.315840  0.445125  0.710
  ↪ 0.4780
Game_Coding_LabelR-N       13.775948 50.246360  0.274
  ↪ 0.7840
Game_Coding_LabelR-Wh       0.429394  0.370369  1.159
  ↪ 0.2463
Game_Coding_LabelR-Y        0.998620  0.540027  1.849
  ↪ 0.0644 .
Game_Coding_LabelReady     -1.392223  0.880250 -1.582
  ↪ 0.1137
Game_Coding_Labelresponse  15.865418 104.597463  0.152
  ↪ 0.8794
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

```

```

# Move type is significant, but no one type significantly
  ↪ differs from any other.

#####
# WHAT IS REFERED TO AND IN WHAT MEDIUM
#####
#####
# EXPLORING ORIENTATION
#####

data_o <- subset(data, O_Speech == "1" | O_Gesture == "1")
data_o_gest <- subset(data_o, Gesture == "1")
data_o_ingest <- subset(data_o, O_Gesture == "1")
data_o_nogest <- subset(data_o, Gesture == "0")

#####
#Is the presence of o in gesture tied to its presence in
  ↪ speech
#####

o_Gest.null <- glmer(O_Gesture ~ 1 + (1+O_Speech|
  ↪ Participant) + (1+O_Speech|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

o_Gest.o_Speech <- glmer(O_Gesture ~ O_Speech + (1+O_Speech
  ↪ |Participant) + (1+O_Speech|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

anova(o_Gest.null, o_Gest.o_Speech)
Data: data
Models:
  o_Gest.null: O_Gesture ~ 1 + (1 + O_Speech | Participant)
    ↪ + (1 + O_Speech | o_Gest.null: Game_Coding_
    ↪ Label)
  o_Gest.o_Speech: O_Gesture ~ O_Speech + (1 + O_Speech |
    ↪ Participant) + (1 + O_Speech | o_Gest.o_Speech:
    ↪ Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
o_Gest.null      7 828.77 868.05 -407.39  814.77
o_Gest.o_Speech  8 823.85 868.74 -403.92  807.85 6.9234
  ↪      1  0.008507 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

```

```

summary(o_Gest.o_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: O_Gesture ~ O_Speech + (1 + O_Speech | Participant
  ↪ ) + (1 + O_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
823.8    868.7   -403.9    807.8     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-0.7969 -0.2552 -0.1768 -0.1213  9.1135

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  1.49148  1.2213
O_Speech         0.01181  0.1087   -1.00
Game_Coding_Label (Intercept)  0.10340  0.3216
O_Speech         0.01588  0.1260   1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.5889      0.4260 -8.425 < 2e-16 ***
O_Speech     0.9119      0.2833  3.218 0.00129 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
O_Speech -0.394

#####
#Is the presence of G in the environment of O negatively
  ↪ correlated with Gesture
#####

o_G_Speech.null <- glmer(G_Speech ~ 1 + (1+Gesture|
  ↪ Participant) + (1+Gesture|Game_Coding_Label), data =
  ↪ data_o, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

o_G_Speech.Gesture <- glmer(G_Speech ~ Gesture + (1+Gesture
  ↪ |Participant) + (1+Gesture|Game_Coding_Label), data =

```



```

→ data_o, family = binomial, control = glmerControl(
→ optimizer = "bobyqa"))

anova(o_G_Speech.null, o_G_Speech.Gesture)
Data: data_o
Models:
  o_G_Speech.null: G_Speech ~ 1 + (1 + Gesture |
    → Participant) + (1 + Gesture | Game_Coding_Label)
  o_G_Speech.Gesture: G_Speech ~ Gesture + (1 + Gesture |
    → Participant) + (1 + Gesture | o_G_Speech.Gesture:
    → Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
o_G_Speech.null      7 690.19 722.02 -338.09   676.19
o_G_Speech.Gesture  8 689.70 726.08 -336.85   673.70 2.4907
→          1    0.1145

summary(o_G_Speech.Gesture)
Generalized linear mixed model fit by maximum likelihood (
→ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ Gesture + (1 + Gesture | Participant) +
→ (1 + Gesture | Game_Coding_Label)
Data: data_o
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
689.7    726.1  -336.8   673.7     690

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.7885  0.3609  0.3951  0.4638  1.7099

Random effects:
  Groups              Name              Variance Std.Dev. Corr
Participant          (Intercept)  0.01441  0.1200
Gesture              0.84949  0.9217   1.00
Game_Coding_Label (Intercept)  0.07093  0.2663
Gesture              0.01692  0.1301  -1.00
Number of obs: 698, groups: Participant, 16; Game_Coding_
→ Label, 11

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.7664      0.1856   9.518 <2e-16 ***
Gesture     -0.7944      0.3871  -2.052  0.0401 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
→ ' ' 1

Correlation of Fixed Effects:

```

```

(Intr)
Gesture -0.232

#####
#Is ground in gesture correlated with orientatin in gesture
  ↪ ?
#####

G_Gesture.null <- glmer(G_Gesture ~ 1 + (1+O_Gesture|
  ↪ Participant) + (1+O_Gesture|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

G_Gesture.O_Gesture <- glmer(G_Gesture ~ O_Gesture + (1+O_
  ↪ Gesture|Participant) + (1+O_Gesture|Game_Coding_Label
  ↪ ), data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(G_Gesture.null, G_Gesture.O_Gesture)
Data: data
Models:
  G_Gesture.null: G_Gesture ~ 1 + (1 + O_Gesture |
    ↪ Participant) + (1 + O_Gesture | G_Gesture.null:
    ↪ Game_Coding_Label)
  G_Gesture.O_Gesture: G_Gesture ~ O_Gesture + (1 + O_Gesture
    ↪ | Participant) + (1 + G_Gesture.O_Gesture: O_
    ↪ Gesture | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Gesture.null      7 197.19 236.47 -91.595   183.19
G_Gesture.O_Gesture  8 175.35 220.24 -79.673   159.35
  ↪ 23.844      1 1.045e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(G_Gesture.O_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Gesture ~ O_Gesture + (1 + O_Gesture |
  ↪ Participant) + (1 + O_Gesture | Game_Coding_
  ↪ Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
175.3    220.2   -79.7    159.3    2013

Scaled residuals:

```

```

  Min      1Q  Median      3Q      Max
-2.4203 -0.0008 -0.0001  0.0000  4.1564

Random effects:
  Groups              Name              Variance Std.Dev. Corr
Participant          (Intercept)  60.12     7.753
O_Gesture              48.00     6.928   -0.98
Game_Coding_Label (Intercept)  49.06     7.005
O_Gesture              48.99     7.000   -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↳ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)   -20.55    13.12  -1.566   0.117
O_Gesture      20.07    13.09   1.534   0.125

Correlation of Fixed Effects:
  (Intr)
O_Gesture -0.999

#####
#Is manner in gesture correlated with orientation?
#####
M_Gesture.null <- glmer(M_Gesture ~ 1 + (1+O_Speech|
  ↳ Participant) + (1|Game_Coding_Label), data = data,
  ↳ family = binomial, control = glmerControl(optimizer =
  ↳ "bobyqa"))

M_Gesture.O_Speech <- glmer(M_Gesture ~ O_Speech + (1+O_
  ↳ Speech|Participant) + (1|Game_Coding_Label), data =
  ↳ data, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))

anova(M_Gesture.null, M_Gesture.O_Speech)
Data: data
Models:
  M_Gesture.null: M_Gesture ~ 1 + (1 + O_Speech |
    ↳ Participant) + (1 | Game_Coding_Label)
  M_Gesture.O_Speech: M_Gesture ~ O_Speech + (1 + O_Speech |
    ↳ Participant) + (1 | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
M_Gesture.null      5 1905.8 1933.8 -947.89   1895.8
M_Gesture.O_Speech  6 1890.5 1924.2 -939.26   1878.5 17.271
  ↳      1 3.241e-05 ***
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
  ↳ ‘ ’ 1

summary(M_Gesture.O_Speech)

```

```

Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: M_Gesture ~ O_Speech + (1 + O_Speech | Participant
  ↪ ) + (1 | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
1890.5   1924.2  -939.3   1878.5    2015

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.9371 -0.5229 -0.3076 -0.1619  5.2333

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  1.36780  1.1695
O_Speech         0.01687  0.1299  -0.97
Game_Coding_Label (Intercept)  0.18784  0.4334
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.1408      0.3449  -3.308  0.00094 ***
O_Speech     -0.8854      0.1477  -5.993  2.07e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
O_Speech -0.279

#####
#Is Direction in gesture correlated with orientation?
#####
Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1+O_Speech|
  ↪ Participant) + (1+O_Speech|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Dir_Gesture.O_Speech <- glmer(Dir_Gesture ~ O_Speech + (1+O
  ↪ _Speech|Participant) + (1+O_Speech|Game_Coding_Label)
  ↪ , data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(Dir_Gesture.null, Dir_Gesture.O_Speech)
Data: data

```

```

Models:
  Dir_Gesture.null: Dir_Gesture ~ 1 + (1 + O_Speech |
    ↪ Participant) + (1 + O_Speech | Dir_Gesture.null:
    ↪ Game_Coding_Label)
  Dir_Gesture.O_Speech: Dir_Gesture ~ O_Speech + (1 + O_
    ↪ Speech | Participant) + (1 + Dir_Gesture.O_Speech: 0
    ↪ _Speech |Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Gesture.null          7 1865.7 1904.9 -925.83  1851.7
Dir_Gesture.O_Speech  8 1855.7 1900.6 -919.85  1839.7
    ↪ 11.956          1  0.0005447 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ '.' 1

summary(Dir_Gesture.O_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Gesture ~ O_Speech + (1 + O_Speech |
  ↪ Participant) + (1 + O_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
1855.7   1900.6   -919.8   1839.7     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.8536 -0.5272 -0.3521 -0.2213  4.7372

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant (Intercept) 1.33015  1.15332
O_Speech      0.16177  0.40220 -0.61
Game_Coding_Label (Intercept) 0.11926  0.34534
O_Speech      0.00728  0.08532  1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.1542      0.3317  -3.48 0.000502 ***
O_Speech     -0.8690      0.2167  -4.01 6.07e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ '.' 1

Correlation of Fixed Effects:
(Intr)

```

```
O_Speech -0.328

#####
#Is orientation in gesture related to manner and direction
  ↪ in gesture.
#####
Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1+O_Gesture|
  ↪ Participant) + (1+O_Gesture|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Dir_Gesture.O_Gesture <- glmer(Dir_Gesture ~ O_Gesture +
  ↪ (1+O_Gesture|Participant) + (1+O_Gesture|Game_Coding_
  ↪ Label), data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(Dir_Gesture.null, Dir_Gesture.O_Gesture)
Data: data
Models:
  Dir_Gesture.null: Dir_Gesture ~ 1 + (1 + O_Gesture |
    ↪ Participant) + (1 + O_Gesture | Dir_Gesture.null:
    ↪ Game_Coding_Label)
  Dir_Gesture.O_Gesture: Dir_Gesture ~ O_Gesture + (1 + O_
    ↪ Gesture | Participant) + (1 + Dir_Gesture.O_Gesture:
    ↪ O_Gesture | Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df
Dir_Gesture.null          7 1869.7 1909.0 -927.83  1855.7
Dir_Gesture.O_Gesture    8 1868.6 1913.5 -926.31  1852.6
  ↪ 3.0513      1
Pr(>Chisq)
Dir_Gesture.null
Dir_Gesture.O_Gesture    0.08067 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(Dir_Gesture.O_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula: Dir_Gesture ~ O_Gesture + (1 + O_Gesture |
  ↪ Participant) + (1 + O_Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
1868.6   1913.5   -926.3   1852.6     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
```

```
-2.7611 -0.5264 -0.3474 -0.2273 3.7707
```

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	1.3676	1.1695	
O_Gesture		2.2015	1.4837	-0.37
Game_Coding_Label	(Intercept)	0.1134	0.3367	
O_Gesture		1.0995	1.0485	1.00

Number of obs: 2021, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.3581	0.3295	-4.121	3.77e-05 ***
O_Gesture	-1.3352	0.8513	-1.568	0.117

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

```
(Intr)
O_Gesture -0.031
```

```
##
```

```
M_Gesture.null <- glmer(M_Gesture ~ 1 + (1+O_Gesture|
  ↳ Participant) + (1+O_Gesture|Game_Coding_Label), data
  ↳ = data, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))
```

```
M_Gesture.O_Gesture <- glmer(M_Gesture ~ O_Gesture + (1+O_
  ↳ Gesture|Participant) + (1+O_Gesture|Game_Coding_Label
  ↳ ), data = data, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))
```

```
anova(M_Gesture.null, M_Gesture.O_Gesture)
```

```
Data: data
```

```
Models:
```

```
M_Gesture.null: M_Gesture ~ 1 + (1 + O_Gesture |
  ↳ Participant) + (1 + O_Gesture | M_Gesture.null:
  ↳ Game_Coding_Label)
M_Gesture.O_Gesture: M_Gesture ~ O_Gesture + (1 + O_Gesture
  ↳ | Participant) + (1 + M_Gesture.O_Gesture: O_
  ↳ Gesture | Game_Coding_Label)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df
M_Gesture.null			7	1926.3	1965.6	-956.17	1912.3
M_Gesture.O_Gesture			8	1924.1	1969.0	-954.05	1908.1
	↳ 4.2287		1				

```
Pr(>Chisq)
```

```
M_Gesture.null
```

```
M_Gesture.O_Gesture 0.03975 *
```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(M_Gesture.O_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula: M_Gesture ~ O_Gesture + (1 + O_Gesture |
  ↪ Participant) + (1 + O_Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
1924.1   1969.0   -954.1   1908.1     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.2162 -0.5265 -0.3305 -0.1298  4.4530

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  1.2360   1.112
O_Gesture        1.5418    1.242    0.18
Game_Coding_Label (Intercept)  0.1452   0.381
O_Gesture        1.4845    1.218    0.86
Number of obs: 2021, groups:  Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3321      0.3220 -4.138 3.51e-05 ***
O_Gesture   -1.5613      0.8758 -1.783  0.0746 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
O_Gesture 0.185

#####
# Do orientation in gesture and direction in gesture predict
  ↪ a reduction of ground in speech?
#####

G_Speech.O_Gesture <- glmer(G_Speech ~ O_Gesture + (1+O_
  ↪ Gesture+M_Gesture|Participant) + (1+O_Gesture+M_
  ↪ Gesture|Game_Coding_Label), data = data, family =

```



```

  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

G_Speech.OM_Gesture <- glmer(G_Speech ~ O_Gesture*M_Gesture
  ↪ + (1+O_Gesture+M_Gesture|Participant) + (1+O_Gesture
  ↪ +M_Gesture|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

anova(G_Speech.O_Gesture, G_Speech.OM_Gesture)
Data: data
Models:
  G_Speech.O_Gesture: G_Speech ~ O_Gesture + (1 + O_Gesture
  ↪ + M_Gesture | Participant) + G_Speech.O_Gesture:
  ↪ (1 + O_Gesture + M_Gesture | Game_Coding_Label)
  G_Speech.OM_Gesture: G_Speech ~ O_Gesture * M_Gesture + (1
  ↪ + O_Gesture + M_Gesture | G_Speech.OM_Gesture:
  ↪ Participant) + (1 + O_Gesture + M_Gesture | Game_
  ↪ Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df
G_Speech.O_Gesture  14 2611.6 2690.2 -1291.8  2583.6
G_Speech.OM_Gesture 16 2607.4 2697.2 -1287.7  2575.4
  ↪ 8.1486      2
Pr(>Chisq)
G_Speech.O_Gesture
G_Speech.OM_Gesture      0.017 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(G_Speech.OM_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula: G_Speech ~ O_Gesture * M_Gesture + (1 + O_Gesture
  ↪ + M_Gesture | Participant) + (1 + O_Gesture + M_
  ↪ Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2607.4  2697.2 -1287.7  2575.4    2005

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.8802 -1.1389  0.6922  0.7852  2.8342

Random effects:
  Groups              Name              Variance Std.Dev. Corr

```

```

Participant      (Intercept) 0.008672 0.09312
O_Gesture    0.269092 0.51874 -0.99
M_Gesture    0.422186 0.64976 -0.60  0.70
Game_Coding_Label (Intercept) 0.046543 0.21574
O_Gesture    0.101334 0.31833  0.42
M_Gesture    0.404878 0.63630  0.74  0.92
Number of obs: 2021, groups:  Participant, 16; Game_Coding_
    ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.51092    0.09974    5.122 3.02e-07 **
    ↪ *
  O_Gesture      0.15601    0.35901    0.435 0.663889
M_Gesture      -1.21827    0.32727   -3.723 0.000197 **
    ↪ *
  O_Gesture:M_Gesture -0.26771    0.62203   -0.430 0.666917
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) O_Gstr M_Gstr
O_Gesture  -0.164
M_Gesture   0.145  0.412
O_Gstr:M_Gs  0.048 -0.308 -0.033

#####
#Exploring Distance
#####

data_dis <- subset(data, Dis_Speech == "1" | Dis_Gesture ==
    ↪ "1")
data_dis_gest <- subset(data_dis, Gesture == "1")
data_dis_ingest <- subset(data_dis, Dis_Gesture == "1")
data_dis_noGest <- subset(data_dis, Gesture == "0")
#####

#####
#Is the presence of orientation in speech correlated with
    ↪ direction in speech?
#####

Dis_Speech.null <- glmer(Dis_Speech ~ 1 + (1+O_Speech|
    ↪ Participant) + (1+O_Speech|Game_Coding_Label), data =
    ↪ data, family = binomial, control = glmerControl(
    ↪ optimizer = "bobyqa"))

Dis_Speech.O_Speech <- glmer(Dis_Speech ~ O_Speech + (1+O_
    ↪ Speech|Participant) + (1+O_Speech|Game_Coding_Label),

```

```

→ data = data, family = binomial, control =
→ glmerControl(optimizer = "bobyqa"))

anova(Dis_Speech.null, Dis_Speech.O_Speech)
Data: data
Models:
  Dis_Speech.null: Dis_Speech ~ 1 + (1 + O_Speech |
    → Participant) + (1 + O_Speech | Dis_Speech.null:
    → Game_Coding_Label)
  Dis_Speech.O_Speech: Dis_Speech ~ O_Speech + (1 + O_Speech
    → | Participant) + (1 + O_Speech | Dis_Speech.O_Speech:
    → Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
Dis_Speech.null      7 1592.5 1631.8 -789.26  1578.5
Dis_Speech.O_Speech  8 1594.4 1639.2 -789.18  1578.4
    → 0.1644      1    0.6851

summary(Dis_Speech.O_Speech)
Generalized linear mixed model fit by maximum likelihood (
  → Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dis_Speech ~ O_Speech + (1 + O_Speech |
  → Participant) + (1 + O_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
1594.4   1639.2  -789.2  1578.4    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-0.9473 -0.4735 -0.3141 -0.1398  6.1339

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant          (Intercept) 1.06481  1.0319
O_Speech             1.49450  1.2225  0.53
Game_Coding_Label (Intercept) 0.06943  0.2635
O_Speech             0.04874  0.2208  -0.65
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  → Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.9351      0.3655 -8.031  9.7e-16 ***
O_Speech     -0.2450      0.6097 -0.402  0.688
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  → ' ' 1

```

Correlation of Fixed Effects:

(Intr)

O\_Speech -0.003

```
#####
#General distribution
#####
#####
#Presence of gesture
#####
data_noGest <- subset(data, Gesture == "0")
data_gest <- subset(data, Gesture == "1")

#####
#Generally, does gesture affect G in speech... yes
#####

G_Speech.null <- glmer(G_Speech ~ 1 + (1+Gesture|
  ↳ Participant) + (1+Gesture|Game_Coding_Label), data =
  ↳ data, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))

G_Speech.Gesture <- glmer(G_Speech ~ Gesture + (1+Gesture|
  ↳ Participant) + (1+Gesture|Game_Coding_Label), data =
  ↳ data, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.Gesture)
Data: data
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + Gesture | Participant)
    ↳ + (1 + Gesture | Game_Coding_Label)
G_Speech.Gesture: G_Speech ~ Gesture + (1 + Gesture |
  ↳ Participant) + (1 + Gesture | G_Speech.Gesture:
  ↳ Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Speech.null      7 2659.1 2698.4 -1322.6    2645.1
G_Speech.Gesture  8 2651.0 2695.9 -1317.5    2635.0 10.155
  ↳      1    0.001439 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

summary(G_Speech.Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↳ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ Gesture + (1 + Gesture | Participant) +
  ↳ (1 + Gesture | Game_Coding_Label)
```

```

Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2651.0   2695.9 -1317.5   2635.0     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.7563 -1.0745  0.6841  0.7937  1.9973

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant (Intercept) 0.01899  0.1378
Gesture     0.27751    0.5268  -0.64
Game_Coding_Label (Intercept) 0.03826  0.1956
Gesture     0.14399    0.3795   0.97
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↳ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.6033     0.1072  5.629 1.81e-08 ***
Gesture     -0.9042     0.2304 -3.925 8.67e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Gesture -0.066

#####
#Generally, does manner and direction in speech affect G in
  ↳ speech
#####

G_Speech.null <- glmer(G_Speech ~ 1 + (1+M_Speech|
  ↳ Participant) + (1+M_Speech|Game_Coding_Label), data =
  ↳ data, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))

G_Speech.M_Speech <- glmer(G_Speech ~ M_Speech + (1+M_
  ↳ Speech|Participant) + (1+M_Speech|Game_Coding_Label),
  ↳ data = data, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.M_Speech)
Data: data
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + M_Speech | Participant

```

```

    ↪ ) + (1 + M_Speech | G_Speech.null:      Game_Coding_
    ↪ Label)
G_Speech.M_Speech: G_Speech ~ M_Speech + (1 + M_Speech |
    ↪ Participant) + (1 + M_Speech | G_Speech.M_Speech:Game
    ↪ _Coding_Label)
Df      AIC      BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
G_Speech.null          7 2634.1 2673.3 -1310.0  2620.1
G_Speech.M_Speech    8 2626.4 2671.3 -1305.2  2610.4 9.698
    ↪          1    0.001845

G_Speech.null
G_Speech.M_Speech **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪  ' ' 1

summary(G_Speech.M_Speech)
Generalized linear mixed model fit by maximum likelihood (
    ↪ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula:
  G_Speech ~ M_Speech + (1 + M_Speech | Participant) + (1 +
    ↪ M_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2626.4    2671.3 -1305.2  2610.4      2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.8116 -0.9882  0.6185  0.7661  2.3316

Random effects:
  Groups          Name      Variance Std.Dev. Corr
Participant (Intercept) 0.129545 0.35992
M_Speech    0.318819 0.56464  0.16
Game_Coding_Label (Intercept) 0.114197 0.33793
M_Speech    0.005477 0.07401 -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
    ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.4607      0.1666  2.766 0.00568 **
M_Speech     -1.0564      0.2158 -4.896 9.76e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪  ' ' 1

```

```

Correlation of Fixed Effects:
  (Intr)
M_Speech -0.187

###
G_Speech.null <- glmer(G_Speech ~ 1 + (1+Dir_Speech|
  ↳ Participant) + (1+Dir_Speech|Game_Coding_Label), data
  ↳ = data, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))

G_Speech.Dir_Speech <- glmer(G_Speech ~ Dir_Speech + (1+Dir
  ↳ _Speech|Participant) + (1+Dir_Speech|Game_Coding_
  ↳ Label), data = data, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.Dir_Speech)
Data: data
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + Dir_Speech |
  ↳ Participant) + (1 + Dir_Speech | G_Speech.null:
  ↳ Game_Coding_Label)
  G_Speech.Dir_Speech: G_Speech ~ Dir_Speech + (1 + Dir_
  ↳ Speech | Participant) + (1 + G_Speech.Dir_Speech:
  ↳ Dir_Speech | Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df
G_Speech.null      7 2737.1 2776.4 -1361.6 2723.1
G_Speech.Dir_Speech 8 2734.9 2779.8 -1359.5 2718.9
  ↳ 4.1982      1
Pr(>Chisq)
G_Speech.null
G_Speech.Dir_Speech 0.04047 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

summary(G_Speech.Dir_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↳ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula: G_Speech ~ Dir_Speech + (1 + Dir_Speech |
  ↳ Participant) + (1 + Dir_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2734.9   2779.8 -1359.5 2718.9    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max

```

-1.5783 -1.0018 0.6597 0.8732 2.1125

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.17778	0.4216	
Dir_Speech	0.12769 0.3573	0.21		
Game_Coding_Label	(Intercept)	0.08960	0.2993	
Dir_Speech	0.08216 0.2866	-0.17		

Number of obs: 2021, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	0.3104	0.1648	1.883	0.0597 .
Dir_Speech	-0.4583	0.2042	-2.244	0.0248 *

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 ↪ ' ' 1

Correlation of Fixed Effects:

(Intr)  
 Dir\_Speech -0.087

```
#####
# Does g in speech predict dir in speech
#####
```

```
Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+G_Speech|
  ↪ Participant) + (1+G_Speech|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))
```

```
Dir_Speech.G_Speech <- glmer(Dir_Speech ~ G_Speech + (1+G_
  ↪ Speech|Participant) + (1+G_Speech|Game_Coding_Label),
  ↪ data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))
```

```
anova(Dir_Speech.null, Dir_Speech.G_Speech)
```

Data: **data**

Models:

```
Dir_Speech.null: Dir_Speech ~ 1 + (1 + G_Speech |
  ↪ Participant) + (1 + G_Speech | Dir_Speech.null:
  ↪ Game_Coding_Label)
Dir_Speech.G_Speech: Dir_Speech ~ G_Speech + (1 + G_Speech
  ↪ | Participant) + (1 + G_Speech | Dir_Speech.G_Speech:
  ↪ Game_Coding_Label)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df
Dir_Speech.null			7	2216.5	2255.8	-1101.3	2202.5
Dir_Speech.G_Speech			8	2216.3	2261.2	-1100.1	2200.3



```

    ↪ 2.2545      1
Pr(>Chisq)
Dir_Speech.null
Dir_Speech.G_Speech      0.1332

summary(Dir_Speech.G_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula:
  Dir_Speech ~ G_Speech + (1 + G_Speech | Participant) + (1
    ↪ + G_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2216.3   2261.2 -1100.1   2200.3     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.4145 -0.6438 -0.4520  0.9614  3.8854

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant          (Intercept) 0.33357  0.5776
G_Speech             0.11530  0.3396  -1.00
Game_Coding_Label (Intercept) 0.46854  0.6845
G_Speech             0.02823  0.1680  -0.34
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3587      0.3031  -4.482 7.38e-06 ***
G_Speech     -0.3726      0.2399  -1.553  0.12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
G_Speech -0.540

#####
# Does g in speech predict man in speech
#####
M_Speech.null <- glmer(M_Speech ~ 1 + (1|Participant) + (1+
  ↪ G_Speech|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

```

```

M_Speech.G_Speech <- glmer(M_Speech ~ G_Speech + (1|
  ↪ Participant) + (1+G_Speech|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

anova(M_Speech.null, M_Speech.G_Speech)
Data: data
Models:
  M_Speech.null: M_Speech ~ 1 + (1 | Participant) + (1 + G_
    ↪ Speech | Game_Coding_Label)
  M_Speech.G_Speech: M_Speech ~ G_Speech + (1 | Participant)
    ↪ + (1 + G_Speech | Game_Coding_Label)
Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
M_Speech.null      5 2291.3 2319.3 -1140.6  2281.3
M_Speech.G_Speech  6 2275.1 2308.8 -1131.6  2263.1 18.152
  ↪      1  2.04e-05

M_Speech.null
M_Speech.G_Speech ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(M_Speech.G_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula:
  M_Speech ~ G_Speech + (1 | Participant) + (1 + G_Speech |
    ↪ Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2275.1  2308.8 -1131.6  2263.1    2015

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.2712 -0.5731 -0.4713  1.0091  4.1383

Random effects:
  Groups              Name              Variance Std.Dev. Corr
Participant          (Intercept)  0.07088  0.2662
Game_Coding_Label (Intercept)  0.43801  0.6618
G_Speech             0.04194  0.2048  0.19
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:

```

```

Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.7071      0.2485  -2.846  0.00443 **
G_Speech    -1.2437      0.2009  -6.191  5.98e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
G_Speech -0.203

#####
#Does g_speech predict gesture?
#####

Gesture.null <- glmer(Gesture ~ 1 + (1+G_Speech|Participant
  ↪ ) + (1+G_Speech|Game_Coding_Label), data = data,
  ↪ family = binomial, control = glmerControl(optimizer =
  ↪ "bobyqa"))

Gesture.G_Speech <- glmer(Gesture ~ G_Speech + (1+G_Speech|
  ↪ Participant) + (1+G_Speech|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

anova(Gesture.null, Gesture.G_Speech)
Data: data
Models:
  Gesture.null: Gesture ~ 1 + (1 + G_Speech | Participant)
  ↪ + (1 + G_Speech | Gesture.null: Game_Coding_Label)
  Gesture.G_Speech: Gesture ~ G_Speech + (1 + G_Speech |
  ↪ Participant) + (1 + G_Speech | Gesture.G_Speech:
  ↪ Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
Gesture.null      7 2233.9 2273.2 -1110.0  2219.9
Gesture.G_Speech  8 2225.1 2270.0 -1104.5  2209.1 10.832
  ↪      1  0.0009979

Gesture.null
Gesture.G_Speech ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

summary(Gesture.G_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula:

```

```

    Gesture ~ G_Speech + (1 + G_Speech | Participant) + (1 +
      ↪ G_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2225.1   2270.0 -1104.5   2209.1     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-4.7830 -0.5846 -0.3507  0.6297  3.1508

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant (Intercept) 2.032795 1.42576
G_Speech    0.372313 0.61017 -0.81
Game_Coding_Label (Intercept) 0.007614 0.08726
G_Speech    0.053134 0.23051  1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.09975    0.37976  0.263    0.793
G_Speech    -0.88587    0.22330 -3.967 7.27e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
G_Speech -0.641

#####
#Does gesture affect the presence of Manner in speech? No
#####

M_Speech.null <- glmer(M_Speech ~ 1 + (1+Gesture|
  ↪ Participant) + (1+Gesture|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

M_Speech.Gesture <- glmer(M_Speech ~ Gesture + (1+Gesture|
  ↪ Participant) + (1+Gesture|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

anova(M_Speech.null, M_Speech.Gesture)
Data: data
Models:

```

```

M_Speech.null: M_Speech ~ 1 + (1 + Gesture | Participant)
  ↪ + (1 + Gesture | Game_Coding_Label)
M_Speech.Gesture: M_Speech ~ Gesture + (1 + Gesture |
  ↪ Participant) + (1 + Gesture | M_Speech.Gesture:
  ↪ Game_Coding_Label)
Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
M_Speech.null      7 2378.1 2417.3 -1182.0  2364.1
M_Speech.Gesture  8 2379.8 2424.7 -1181.9  2363.8 0.2781
  ↪          1      0.5979

summary(M_Speech.Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: M_Speech ~ Gesture + (1 + Gesture | Participant) +
  ↪ (1 + Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2379.8    2424.7 -1181.9  2363.8    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.0970 -0.6687 -0.5176  1.0959  3.3753

Random effects:
  Groups              Name              Variance Std.Dev. Corr
Participant (Intercept) 0.1681    0.4100
Gesture      0.5743    0.7578   -1.00
Game_Coding_Label (Intercept) 0.4534    0.6734
Gesture      0.1679    0.4098   -0.16
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.4420    0.2703  -5.336 9.51e-08 ***
Gesture      0.1656    0.3095   0.535  0.592
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Gesture -0.520

#####
#Does gesture affect the presence of Direction in speech?

```

```

    ↪ Yes
#####
Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+Gesture|
    ↪ Participant) + (1+Gesture|Game_Coding_Label), data =
    ↪ data, family = binomial, control = glmerControl(
    ↪ optimizer = "bobyqa"))

Dir_Speech.Gesture <- glmer(Dir_Speech ~ Gesture + (1+
    ↪ Gesture|Participant) + (1+Gesture|Game_Coding_Label),
    ↪ data = data, family = binomial, control =
    ↪ glmerControl(optimizer = "bobyqa"))

anova(Dir_Speech.null, Dir_Speech.Gesture)
Data: data
Models:
  Dir_Speech.null: Dir_Speech ~ 1 + (1 + Gesture |
    ↪ Participant) + (1 + Gesture | Dir_Speech.null:
    ↪ Game_Coding_Label)
  Dir_Speech.Gesture: Dir_Speech ~ Gesture + (1 + Gesture |
    ↪ Participant) + (1 + Gesture | Dir_Speech.Gesture:Game
    ↪ _Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Speech.null      7 2233.6 2272.9 -1109.8   2219.6
Dir_Speech.Gesture  8 2228.8 2273.7 -1106.4   2212.8 6.7629
    ↪      1    0.009307 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪  ' ' 1

summary(Dir_Speech.Gesture)
Generalized linear mixed model fit by maximum likelihood (
    ↪ Laplace Approximation) ['glmerMod']
Family: binomial (logit)
Formula: Dir_Speech ~ Gesture + (1 + Gesture | Participant)
    ↪ + (1 + Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2228.8   2273.7  -1106.4   2212.8    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.0601 -0.6287 -0.4400  0.9433  3.7799

Random effects:
  Groups             Name             Variance Std.Dev. Corr
Participant          (Intercept)  0.22283  0.4720

```

```

Gesture      0.08271  0.2876  -0.75
Game_Coding_Label (Intercept) 0.63386  0.7962
Gesture      0.04161  0.2040  -1.00
Number of obs: 2021, groups:  Participant, 16; Game_Coding_
  ↳ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.8418      0.3196  -5.764 8.24e-09 ***
Gesture      0.5503      0.2157   2.551 0.0107 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ '.' 1

Correlation of Fixed Effects:
  (Intr)
Gesture -0.642

#####
#Manner in gesture by manner in speech?
#####

M_Gesture.null <- glmer(M_Gesture ~ 1 + (1+M_Speech|
  ↳ Participant) + (1+M_Speech|Game_Coding_Label), data =
  ↳ data, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))

M_Gesture.M_Speech <- glmer(M_Gesture ~ M_Speech + (1+M_
  ↳ Speech|Participant) + (1+M_Speech|Game_Coding_Label),
  ↳ data = data, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))

anova(M_Gesture.null, M_Gesture.M_Speech)
Data: data
Models:
  M_Gesture.null: M_Gesture ~ 1 + (1 + M_Speech |
    ↳ Participant) + (1 + M_Speech | M_Gesture.null: Game
    ↳ _Coding_Label)
  M_Gesture.M_Speech: M_Gesture ~ M_Speech + (1 + M_Speech |
    ↳ Participant) + (1 + M_Speech |M_Gesture.M_Speech:Game
    ↳ _Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
M_Gesture.null      7 1864.3 1903.6 -925.14  1850.3
M_Gesture.M_Speech  8 1852.9 1897.8 -918.45  1836.9 13.393
  ↳      1 0.0002525 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ '.' 1

```

```

summary(M_Gesture.M_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: M_Gesture ~ M_Speech + (1 + M_Speech | Participant
  ↪ ) + (1 + M_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
1852.9   1897.8   -918.4   1836.9     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.0635 -0.5275 -0.2781 -0.1536  4.9134

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  1.3127   1.1457
M_Speech         0.3860    0.6213  -0.24
Game_Coding_Label (Intercept)  0.2408   0.4907
M_Speech         0.1289    0.3590  -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.7396     0.3531 -4.927 8.36e-07 ***
M_Speech     1.2036     0.2616  4.602 4.19e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
M_Speech -0.457

#####
#Direction in gesture by manner in speech?
#####

Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1+M_Speech|
  ↪ Participant) + (1+M_Speech|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Dir_Gesture.M_Speech <- glmer(Dir_Gesture ~ M_Speech + (1+M
  ↪ _Speech|Participant) + (1+M_Speech|Game_Coding_Label)
  ↪ , data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

```



```

anova(Dir_Gesture.null,Dir_Gesture.M_Speech)
Data: data
Models:
  Dir_Gesture.null: Dir_Gesture ~ 1 + (1 + M_Speech |
    ↪ Participant) + (1 + M_Speech | Dir_Gesture.null:
    ↪ Game_Coding_Label)
  Dir_Gesture.M_Speech: Dir_Gesture ~ M_Speech + (1 + M_
    ↪ Speech | Participant) + (1 + Dir_Gesture.M_Speech:
    ↪ M_Speech | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Gesture.null      7 1851.9 1891.2 -918.96  1837.9
Dir_Gesture.M_Speech  8 1848.8 1893.7 -916.40  1832.8
    ↪ 5.1231      1   0.02361 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

summary(Dir_Gesture.M_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Gesture ~ M_Speech + (1 + M_Speech |
  ↪ Participant) + (1 +      M_Speech | Game_Coding_Label
  ↪ )
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
1848.8   1893.7  -916.4  1832.8    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.3294 -0.5160 -0.3339 -0.2041  4.6357

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  1.16859  1.0810
M_Speech         0.25171  0.5017  -0.32
Game_Coding_Label (Intercept)  0.13920  0.3731
M_Speech         0.02909  0.1706  -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.6163      0.3205  -5.043 4.57e-07 ***
M_Speech     0.7925      0.2151   3.684 0.000229 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

```

↪ ‘ ’ 1

Correlation of Fixed Effects:

(Intr)

M\_Speech -0.390

```
#####
#Manner in gesture by direction in speech?
#####
```

```
M_Gesture.null <- glmer(M_Gesture ~ 1 + (1+Dir_Speech|
  ↪ Participant) + (1+Dir_Speech|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))
```

```
M_Gesture.Dir_Speech <- glmer(M_Gesture ~ Dir_Speech + (1+
  ↪ Dir_Speech|Participant) + (1+Dir_Speech|Game_Coding_
  ↪ Label), data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))
```

**anova**(M\_Gesture.null, M\_Gesture.Dir\_Speech)

Data: **data**

Models:

```
M_Gesture.null: M_Gesture ~ 1 + (1 + Dir_Speech |
  ↪ Participant) + (1 + Dir_Speech | M_Gesture.null:
  ↪ Game_Coding_Label)
```

```
M_Gesture.Dir_Speech: M_Gesture ~ Dir_Speech + (1 + Dir_
  ↪ Speech | Participant) + (1 + M_Gesture.Dir_Speech:
  ↪ Dir_Speech |Game_Coding_Label)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
M_Gesture.null			7	1897.7	1937	-941.84		1883.7
M_Gesture.Dir_Speech			8	1891.1	1936	-937.55		1875.1 8.5958
			↪ 1	0.003369	**			

---

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1
```

**summary**(M\_Gesture.Dir\_Speech)

Generalized linear mixed **model** fit **by** maximum likelihood (
 ↪ Laplace Approximation) ['glmerMod']

Family: **binomial** ( logit )

```
Formula: M_Gesture ~ Dir_Speech + (1 + Dir_Speech |
  ↪ Participant) + (1 + Dir_Speech | Game_Coding_
  ↪ Label)
```

Data: **data**

Control: glmerControl(optimizer = "bobyqa")

AIC	BIC	logLik	deviance	df.resid
-----	-----	--------	----------	----------

```

1891.1   1936.0   -937.5   1875.1   2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.7081 -0.5249 -0.3094 -0.1488  4.9205

Random effects:
  Groups              Name              Variance Std.Dev. Corr
Participant          (Intercept)  1.22382  1.1063
Dir_Speech  0.02394  0.1547   0.42
Game_Coding_Label (Intercept)  0.31408  0.5604
Dir_Speech  0.39172  0.6259  -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.6316      0.3513  -4.645 3.41e-06 ***
Dir_Speech   1.1117      0.2646   4.201 2.65e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Dir_Speech -0.442

#####
#Direction in gesture by direction in speech?
#####

Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1+Dir_Speech|
  ↪ Participant) + (1+Dir_Speech|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Dir_Gesture.Dir_Speech <- glmer(Dir_Gesture ~ Dir_Speech +
  ↪ (1+Dir_Speech|Participant) + (1+Dir_Speech|Game_
  ↪ Coding_Label), data = data, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

anova(Dir_Gesture.null, Dir_Gesture.Dir_Speech)
Data: data
Models:
  Dir_Gesture.null: Dir_Gesture ~ 1 + (1 + Dir_Speech |
  ↪ Participant) + (1 + Dir_Speech | Dir_Gesture.null:
  ↪ Game_Coding_Label)
Dir_Gesture.Dir_Speech: Dir_Gesture ~ Dir_Speech + (1 + Dir
  ↪ _Speech | Participant) + (1 + Dir_Gesture.Dir_Speech:
  ↪ Dir_Speech | Game_Coding_Label)

```

```

Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Gesture.null          7 1812.7 1852.0 -899.35  1798.7
Dir_Gesture.Dir_Speech  8 1803.0 1847.9 -893.52  1787.0
  ↪ 11.647          1  0.0006429 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(Dir_Gesture.Dir_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Gesture ~ Dir_Speech + (1 + Dir_Speech |
  ↪ Participant) + (1 + Dir_Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
1803.0    1847.9   -893.5   1787.0     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.7568 -0.4714 -0.3205 -0.2042  4.3080

Random effects:
  Groups          Name      Variance Std.Dev. Corr
Participant      (Intercept) 0.9738  0.9868
Dir_Speech 0.1228  0.3504  1.00
Game_Coding_Label (Intercept) 0.3281  0.5728
Dir_Speech 0.5942  0.7708 -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.7326      0.3292 -5.263 1.42e-07 ***
Dir_Speech  1.4172      0.3121  4.541 5.61e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Dir_Speech -0.318

#####
#Manner in Speech by manner in Gesture?
#####

```

```

M_Speech.null <- glmer(M_Speech ~ 1 + (1+M_Gesture|
  ↪ Participant) + (1+M_Gesture|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

M_Speech.M_Gesture <- glmer(M_Speech ~ M_Gesture + (1+M_
  ↪ Gesture|Participant) + (1+M_Gesture|Game_Coding_Label
  ↪ ), data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(M_Speech.null, M_Speech.M_Gesture)
Data: data
Models:
  M_Speech.null: M_Speech ~ 1 + (1 + M_Gesture |
    ↪ Participant) + (1 + M_Gesture | M_Speech.null: Game
    ↪ _Coding_Label)
  M_Speech.M_Gesture: M_Speech ~ M_Gesture + (1 + M_Gesture |
    ↪ Participant) + (1 + M_Gesture | M_Speech.M_Gesture:
    ↪ Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
M_Speech.null      7 2325.3 2364.6 -1155.6 2311.3
M_Speech.M_Gesture 8 2318.2 2363.1 -1151.1 2302.2 9.1228
  ↪      1 0.002524 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(M_Speech.M_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: M_Speech ~ M_Gesture + (1 + M_Gesture |
  ↪ Participant) + (1 + M_Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2318.2   2363.1 -1151.1 2302.2    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.4669 -0.6304 -0.5278  0.9700  3.6790

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  0.1426   0.3777
M_Gesture         0.6419   0.8012  -0.91
Game_Coding_Label (Intercept)  0.4880   0.6986
M_Gesture         0.2027   0.4502  -0.76
Number of obs: 2021, groups: Participant, 16; Game_Coding_

```

↪ Label, 13

Fixed **effects**:

```
Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.6414      0.2722  -6.030 1.64e-09 ***
M_Gesture    1.1560      0.3386   3.414 0.000639 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1
```

Correlation of Fixed Effects:

```
(Intr)
M_Gesture -0.638
```

```
#####
#Direction in Speech by manner in Gesture?
#####
```

```
Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+M_Gesture|
  ↪ Participant) + (1+M_Gesture|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))
```

```
Dir_Speech.M_Gesture <- glmer(Dir_Speech ~ M_Gesture + (1+M
  ↪ _Gesture|Participant) + (1+M_Gesture|Game_Coding_
  ↪ Label), data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))
```

```
anova(Dir_Speech.null, Dir_Speech.M_Gesture)
```

```
Data: data
```

```
Models:
```

```
Dir_Speech.null: Dir_Speech ~ 1 + (1 + M_Gesture |
  ↪ Participant) + (1 + M_Gesture | Dir_Speech.null:
  ↪ Game_Coding_Label)
```

```
Dir_Speech.M_Gesture: Dir_Speech ~ M_Gesture + (1 + M_
  ↪ Gesture | Participant) + (1 + Dir_Speech.M_Gesture: M
  ↪ _Gesture | Game_Coding_Label)
```

```
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
Dir_Speech.null      7 2206.2 2245.4 -1096.1 2192.2
Dir_Speech.M_Gesture  8 2197.8 2242.7 -1090.9 2181.8
  ↪ 10.361      1 0.001287 **
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1
```

```
summary(Dir_Speech.M_Gesture)
```

```
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
```

```

Family: binomial ( logit )
Formula: Dir_Speech ~ M_Gesture + (1 + M_Gesture |
  ↳ Participant) + (1 + M_Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2197.8   2242.7 -1090.9  2181.8    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.2346 -0.6854 -0.4269  0.9912  4.2077

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  0.21865  0.4676
M_Gesture        0.02939  0.1714   -0.36
Game_Coding_Label (Intercept)  0.55920  0.7478
M_Gesture        0.18110  0.4256   -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↳ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.8686     0.2960 -6.313 2.74e-10 ***
M_Gesture    1.1554     0.2418  4.779 1.77e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

Correlation of Fixed Effects:
  (Intr)
M_Gesture -0.665

#####
#Manner in Speech by Direction in Gesture? No. This
  ↳ suggests that gesture depicts direction when speech
  ↳ is not describing manner
#####

M_Speech.null <- glmer(M_Speech ~ 1 + (1+Dir_Gesture|
  ↳ Participant) + (1+Dir_Gesture|Game_Coding_Label),
  ↳ data = data, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))

M_Speech.Dir_Gesture <- glmer(M_Speech ~ Dir_Gesture + (1+
  ↳ Dir_Gesture|Participant) + (1+Dir_Gesture|Game_Coding
  ↳ _Label), data = data, family = binomial, control =

```

```
↪ glmerControl(optimizer = "bobyqa"))
```

```
anova(M_Speech.null, M_Speech.Dir_Gesture)
```

```
Data: data
```

```
Models:
```

```
M_Speech.null: M_Speech ~ 1 + (1 + Dir_Gesture |
  ↪ Participant) + (1 + Dir_Gesture | M_Speech.null:
  ↪ Game_Coding_Label)
```

```
M_Speech.Dir_Gesture: M_Speech ~ Dir_Gesture + (1 + Dir_
  ↪ Gesture | Participant) + (1 + M_Speech.Dir_Gesture:
  ↪ Dir_Gesture | Game_Coding_Label)
```

Df	AIC	BIC	logLik	<b>deviance</b>	Chisq	Chi Df	Pr(>Chisq)
M_Speech.null			7 2353.1	2392.4	-1169.6		2339.1
M_Speech.Dir_Gesture			8 2352.2	2397.1	-1168.1		2336.2
	↪ 2.9423	1	0.08629	.			

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1
```

```
summary(M_Speech.Dir_Gesture)
```

```
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
```

```
Family: binomial ( logit )
```

```
Formula: M_Speech ~ Dir_Gesture + (1 + Dir_Gesture |
  ↪ Participant) + (1 + Dir_Gesture | Game_Coding_Label)
```

```
Data: data
```

```
Control: glmerControl(optimizer = "bobyqa")
```

AIC	BIC	logLik	<b>deviance</b>	<b>df.resid</b>
2352.2	2397.1	-1168.1	2336.2	2013

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-1.2927	-0.6570	-0.5378	0.9696	3.6417

```
Random effects:
```

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.1005	0.3171	
Dir_Gesture		0.5690	0.7543	-1.00
Game_Coding_Label	(Intercept)	0.4797	0.6926	
Dir_Gesture		0.2680	0.5177	-0.43

```
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13
```

```
Fixed effects:
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.5162	0.2593	-5.847	4.99e-09 ***
Dir_Gesture	0.6894	0.3576	1.928	0.0539 .

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```



```

    ↪ ‘ ’ 1

Correlation of Fixed Effects:
  (Intr)
Dir_Gesture -0.498

#####
#Direction in Speech by Direction in Gesture?
#####

Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+Dir_Gesture|
  ↪ Participant) + (1+Dir_Gesture|Game_Coding_Label),
  ↪ data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

Dir_Speech.Dir_Gesture <- glmer(Dir_Speech ~ Dir_Gesture +
  ↪ (1+Dir_Gesture|Participant) + (1+Dir_Gesture|Game_
  ↪ Coding_Label), data = data, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

anova(Dir_Speech.null, Dir_Speech.Dir_Gesture)
Data: data
Models:
  Dir_Speech.null: Dir_Speech ~ 1 + (1 + Dir_Gesture |
    ↪ Participant) + (1 + Dir_Gesture | Dir_Speech.null:
    ↪ Game_Coding_Label)
  Dir_Speech.Dir_Gesture: Dir_Speech ~ Dir_Gesture + (1 + Dir
    ↪ _Gesture | Participant) + Dir_Speech.Dir_Gesture:
    ↪ (1 + Dir_Gesture | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Speech.null          7 2169.7 2209.0 -1077.9  2155.7
Dir_Speech.Dir_Gesture  8 2158.8 2203.7 -1071.4  2142.8
  ↪ 12.959      1  0.0003183 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
  ↪ ‘ ’ 1

summary(Dir_Speech.Dir_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Speech ~ Dir_Gesture + (1 + Dir_Gesture |
  ↪ Participant) + (1 + Dir_Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid

```

2158.8    2203.6    -1071.4    2142.8    2013

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-1.3100	-0.6725	-0.4175	0.9513	4.7390

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.26807	0.5178	
Dir_Gesture		0.04389	0.2095	-0.86
Game_Coding_Label	(Intercept)	0.65475	0.8092	
Dir_Gesture		0.45576	0.6751	-0.91

Number of obs: 2021, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.0322	0.3256	-6.241 4.34e-10 ***
Dir_Gesture	1.5963	0.3086	5.173 2.31e-07 ***

---  
 Signif. **codes**: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 ↪ ' ' 1

Correlation of Fixed Effects:

(Intr)  
 Dir\_Gesture -0.750

```
#####
#Manner in Speech by Direction in Speech?
#####
```

```
M_Speech.null <- glmer(M_Speech ~ 1 + (1+Dir_Speech|
  ↪ Participant) + (1+Dir_Speech|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))
```

```
M_Speech.Dir_Speech <- glmer(M_Speech ~ Dir_Speech + (1+Dir
  ↪ _Speech|Participant) + (1+Dir_Speech|Game_Coding_
  ↪ Label), data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))
```

```
anova(M_Speech.null, M_Speech.Dir_Speech)
```

Data: **data**

Models:

```
M_Speech.null: M_Speech ~ 1 + (1 + Dir_Speech |
  ↪ Participant) + (1 + Dir_Speech | M_Speech.null:
  ↪ Game_Coding_Label)
M_Speech.Dir_Speech: M_Speech ~ Dir_Speech + (1 + Dir_
  ↪ Speech | Participant) + (1 + M_Speech.Dir_Speech:
```

```

    ↪ Dir_Speech | Game_Coding_Label)
Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
M_Speech.null          7 2395.6 2434.9 -1190.8   2381.6
M_Speech.Dir_Speech  8 2397.4 2442.3 -1190.7   2381.4
    ↪ 0.1905          1      0.6625

#####
#Direction in Speech by Manner in Speech?
#####
Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+M_Speech|
    ↪ Participant) + (1+M_Speech|Game_Coding_Label), data =
    ↪ data, family = binomial, control = glmerControl(
    ↪ optimizer = "bobyqa"))

Dir_Speech.M_Speech <- glmer(Dir_Speech ~ M_Speech + (1+M_
    ↪ Speech|Participant) + (1+M_Speech|Game_Coding_Label),
    ↪ data = data, family = binomial, control =
    ↪ glmerControl(optimizer = "bobyqa"))

anova(Dir_Speech.null, Dir_Speech.M_Speech)
Data: data
Models:
  Dir_Speech.null: Dir_Speech ~ 1 + (1 + M_Speech |
    ↪ Participant) + (1 + M_Speech | Dir_Speech.null:Game
    ↪ _Coding_Label)
  Dir_Speech.M_Speech: Dir_Speech ~ M_Speech + (1 + M_Speech
    ↪ | Participant) + (1 + M_Speech | Dir_Speech.M_Speech:
    ↪ Game_Coding_Label)
Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Speech.null          7 2225.7 2265.0 -1105.8   2211.7
Dir_Speech.M_Speech  8 2226.8 2271.7 -1105.4   2210.8
    ↪ 0.8556          1      0.355

summary(Dir_Speech.M_Speech)
Generalized linear mixed model fit by maximum likelihood (
    ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Speech ~ M_Speech + (1 + M_Speech |
    ↪ Participant) + (1 + M_Speech |      Game_Coding_Label
    ↪ )
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2226.8    2271.7 -1105.4   2210.8      2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.1899 -0.6208 -0.4509  0.8404  4.0295

```

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.161046	0.40131	
M_Speech	0.008017 0.08954	-1.00		
Game_Coding_Label	(Intercept)	0.634998	0.79687	
M_Speech	0.240275 0.49018	-1.00		

Number of obs: 2021, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.5423	0.2936	-5.252	1.5e-07 ***
M_Speech	0.2095	0.2410	0.869	0.385

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)  
 M\_Speech -0.697

```
#####
#Manner in Gesture by Direction in Gesture?
#####
```

```
M_Gesture.null <- glmer(M_Gesture ~ 1 + (1+Dir_Gesture|
  ↳ Participant) + (1+Dir_Gesture|Game_Coding_Label),
  ↳ data = data, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))
```

```
M_Gesture.Dir_Gesture <- glmer(M_Gesture ~ Dir_Gesture +
  ↳ (1+Dir_Gesture|Participant) + (1+Dir_Gesture|Game_
  ↳ Coding_Label), data = data, family = binomial,
  ↳ control = glmerControl(optimizer = "bobyqa"))
```

```
anova(M_Gesture.null, M_Gesture.Dir_Gesture)
```

Data: **data**

Models:

```
M_Gesture.null: M_Gesture ~ 1 + (1 + Dir_Gesture |
  ↳ Participant) + (1 + Dir_Gesture | M_Gesture.null:
  ↳ Game_Coding_Label)
M_Gesture.Dir_Gesture: M_Gesture ~ Dir_Gesture + (1 + Dir_
  ↳ Gesture | Participant) + (1 + M_Gesture.Dir_Gesture:
  ↳ Dir_Gesture | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
M_Gesture.null          7 922.08 961.36 -454.04  908.08
M_Gesture.Dir_Gesture  8 890.60 935.50 -437.30  874.60
  ↳ 33.473      1 7.226e-09 ***
---
```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                ↪ ' ' 1

summary(M_Gesture.Dir_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: M_Gesture ~ Dir_Gesture + (1 + Dir_Gesture |
  ↪ Participant) + (1 + Dir_Gesture | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
890.6    935.5  -437.3   874.6    2013

Scaled residuals:
  Min       1Q   Median       3Q      Max
-6.8906 -0.1567 -0.1187 -0.0683  9.7828

Random effects:
  Groups             Name                Variance Std.Dev. Corr
Participant          (Intercept)  2.297106  1.51562
Dir_Gesture  2.257383  1.50246  -0.86
Game_Coding_Label (Intercept)  0.153538  0.39184
Dir_Gesture  0.007159  0.08461  -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.7076      0.5216  -7.108 1.18e-12 ***
Dir_Gesture  6.0321      0.5897  10.229 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Dir_Gesture -0.778

#####
#Direction in Gesture by Manner in Gesture?
#####

Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1+M_Gesture|
  ↪ Participant) + (1+M_Gesture|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Dir_Gesture.M_Gesture <- glmer(Dir_Gesture ~ M_Gesture +

```

```

↪ (1+M_Gesture|Participant) + (1+M_Gesture|Game_Coding_
↪ Label), data = data, family = binomial, control =
↪ glmerControl(optimizer = "bobyqa"))

```

```
anova(Dir_Gesture.null, Dir_Gesture.M_Gesture)
```

```
Data: data
```

```
Models:
```

```
Dir_Gesture.null: Dir_Gesture ~ 1 + (1 + M_Gesture |
↪ Participant) + (1 + M_Gesture | Dir_Gesture.null:
↪ Game_Coding_Label)
```

```
Dir_Gesture.M_Gesture: Dir_Gesture ~ M_Gesture + (1 + M_
↪ Gesture | Participant) + (1 + Dir_Gesture.M_Gesture:
↪ M_Gesture | Game_Coding_Label)
```

```
Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
Dir_Gesture.null 7 874.93 914.21 -430.47 860.93
```

```
Dir_Gesture.M_Gesture 8 839.45 884.34 -411.73 823.45
```

```
↪ 37.482 1 9.225e-10 ***
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
↪ ' ' 1
```

```
summary(Dir_Gesture.M_Gesture)
```

```
Generalized linear mixed model fit by maximum likelihood (
↪ Laplace Approximation) ['glmerMod']
```

```
Family: binomial ( logit )
```

```
Formula: Dir_Gesture ~ M_Gesture + (1 + M_Gesture |
↪ Participant) + (1 + M_Gesture | Game_Coding_
↪ Label)
```

```
Data: data
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```
AIC BIC logLik deviance df.resid
839.5 884.3 -411.7 823.5 2013
```

```
Scaled residuals:
```

```
Min 1Q Median 3Q Max
-3.5707 -0.1801 -0.1181 -0.0716 12.2004
```

```
Random effects:
```

```
Groups Name Variance Std.Dev. Corr
Participant (Intercept) 1.663e+00 1.290e+00
M_Gesture 1.749e+00 1.322e+00 -0.84
Game_Coding_Label (Intercept) 0.000e+00 0.000e+00
```

```
M_Gesture 1.861e-11 4.314e-06 NaN
```

```
Number of obs: 2021, groups: Participant, 16; Game_Coding_
↪ Label, 13
```

```
Fixed effects:
```

```
Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept) -3.9166      0.4537  -8.633   <2e-16 ***
M_Gesture   5.8773      0.5299  11.091   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
M_Gesture -0.809

#####
#Manner in Speech by Ground in Speech?
#####

M_Speech.null <- glmer(M_Speech ~ 1 + (1|Participant) + (1+
  ↪ G_Speech|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

M_Speech.G_Speech <- glmer(M_Speech ~ G_Speech + (1|
  ↪ Participant) + (1+G_Speech|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

anova(M_Speech.null, M_Speech.G_Speech)
Data: data
Models:
  M_Speech.null: M_Speech ~ 1 + (1 | Participant) + (1 + G_
  ↪ Speech | Game_Coding_Label)
  M_Speech.G_Speech: M_Speech ~ G_Speech + (1 | Participant)
  ↪ + (1 + G_Speech | Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
M_Speech.null      5 2291.3 2319.3 -1140.6  2281.3
M_Speech.G_Speech  6 2275.1 2308.8 -1131.6  2263.1 18.152
  ↪      1  2.04e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

summary(M_Speech.G_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: M_Speech ~ G_Speech + (1 | Participant) + (1 + G_
  ↪ Speech | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
```

2275.1 2308.8 -1131.6 2263.1 2015

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-1.2712	-0.5731	-0.4713	1.0091	4.1383

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.07088	0.2662	
Game_Coding_Label	(Intercept)	0.43801	0.6618	
G_Speech	0.04194	0.2048	0.19	

Number of obs: 2021, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.7071	0.2485	-2.846 0.00443 **
G_Speech	-1.2437	0.2009	-6.191 5.98e-10 ***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 ↪ ' ' 1

Correlation of Fixed Effects:

(Intr)  
 G\_Speech -0.203

```
#####
#Direction in Speech by Ground in Speech?
#####
```

```
Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+G_Speech|
  ↪ Participant) + (1+G_Speech|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))
```

```
Dir_Speech.G_Speech <- glmer(Dir_Speech ~ G_Speech + (1+G_
  ↪ Speech|Participant) + (1+G_Speech|Game_Coding_Label),
  ↪ data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))
```

```
anova(Dir_Speech.null, Dir_Speech.G_Speech)
```

```
Data: data
```

```
Models:
```

```
Dir_Speech.null: Dir_Speech ~ 1 + (1 + G_Speech |
  ↪ Participant) + (1 + G_Speech | Dir_Speech.null:
  ↪ Game_Coding_Label)
Dir_Speech.G_Speech: Dir_Speech ~ G_Speech + (1 + G_Speech
  ↪ | Participant) + (1 + G_Speech | Dir_Speech.G_Speech:
  ↪ Game_Coding_Label)
```



```

Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Speech.null      7 2216.5 2255.8 -1101.3  2202.5
Dir_Speech.G_Speech  8 2216.3 2261.2 -1100.1  2200.3
  ↳ 2.2545      1      0.1332

summary(Dir_Speech.G_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↳ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Speech ~ G_Speech + (1 + G_Speech |
  ↳ Participant) + (1 + G_Speech |      Game_Coding_Label
  ↳ )
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2216.3    2261.2 -1100.1  2200.3    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.4145 -0.6438 -0.4520  0.9614  3.8854

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  0.33357  0.5776
G_Speech         0.11530  0.3396  -1.00
Game_Coding_Label (Intercept)  0.46854  0.6845
G_Speech         0.02823  0.1680  -0.34
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↳ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3587      0.3031  -4.482 7.38e-06 ***
G_Speech    -0.3726      0.2399  -1.553  0.12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

Correlation of Fixed Effects:
  (Intr)
G_Speech -0.540

#####
#Ground in Speech by Manner in Speech?
#####

G_Speech.null <- glmer(G_Speech ~ 1 + (1+M_Speech|
  ↳ Participant) + (1+M_Speech|Game_Coding_Label), data =

```

```

↪ data, family = binomial, control = glmerControl(
↪ optimizer = "bobyqa"))

```

```

G_Speech.M_Speech <- glmer(G_Speech ~ M_Speech + (1+M_
↪ Speech|Participant) + (1+M_Speech|Game_Coding_Label),
↪ data = data, family = binomial, control =
↪ glmerControl(optimizer = "bobyqa"))

```

```
anova(G_Speech.null, G_Speech.M_Speech)
```

```
Data: data
```

```
Models:
```

```

G_Speech.null: G_Speech ~ 1 + (1 + M_Speech | Participant
↪ ) + (1 + M_Speech | G_Speech.null: Game_Coding_
↪ Label)

```

```

G_Speech.M_Speech: G_Speech ~ M_Speech + (1 + M_Speech |
↪ Participant) + (1 + M_Speech | G_Speech.M_Speech:
↪ Game_Coding_Label)

```

```

Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
G_Speech.null      7 2634.1 2673.3 -1310.0  2620.1
G_Speech.M_Speech  8 2626.4 2671.3 -1305.2  2610.4 9.698
↪      1    0.001845 **

```

```
---
```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
↪ ' ' 1

```

```
summary(G_Speech.M_Speech)
```

```

Generalized linear mixed model fit by maximum likelihood (
↪ Laplace Approximation) ['glmerMod']

```

```
Family: binomial (logit)
```

```

Formula: G_Speech ~ M_Speech + (1 + M_Speech | Participant)
↪ + (1 + M_Speech | Game_Coding_Label)

```

```
Data: data
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```

AIC      BIC    logLik deviance df.resid
2626.4   2671.3  -1305.2  2610.4    2013

```

```
Scaled residuals:
```

```

  Min      1Q  Median      3Q      Max
-1.8116 -0.9882  0.6185  0.7661  2.3316

```

```
Random effects:
```

```

  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  0.129545 0.35992
M_Speech         0.318819 0.56464  0.16
Game_Coding_Label (Intercept)  0.114197 0.33793
M_Speech         0.005477 0.07401  -1.00

```

```

Number of obs: 2021, groups: Participant, 16; Game_Coding_
↪ Label, 13

```

```

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.4607      0.1666  2.766  0.00568 **
M_Speech     -1.0564      0.2158 -4.896  9.76e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
M_Speech -0.187

#####
#Ground in Speech by Direction in Speech?
#####

G_Speech.null <- glmer(G_Speech ~ 1 + (1+Dir_Speech|
  ↪ Participant) + (1+Dir_Speech|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

G_Speech.Dir_Speech <- glmer(G_Speech ~ Dir_Speech + (1+Dir
  ↪ _Speech|Participant) + (1+Dir_Speech|Game_Coding_
  ↪ Label), data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.Dir_Speech)
Data: data
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + Dir_Speech |
  ↪ Participant) + (1 + Dir_Speech | G_Speech.null:
  ↪ Game_Coding_Label)
  G_Speech.Dir_Speech: G_Speech ~ Dir_Speech + (1 + Dir_
  ↪ Speech | Participant) + (1 + G_Speech.Dir_Speech:
  ↪ Dir_Speech | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Speech.null          7 2737.1 2776.4 -1361.6 2723.1
G_Speech.Dir_Speech  8 2734.9 2779.8 -1359.5 2718.9
  ↪ 4.1982      1 0.04047 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

summary(G_Speech.Dir_Speech)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ Dir_Speech + (1 + Dir_Speech |
  ↪ Participant) + (1 + Dir_Speech | Game_Coding_Label)

```

```

Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2734.9   2779.8  -1359.5   2718.9     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.5783 -1.0018  0.6597  0.8732  2.1125

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant          (Intercept) 0.17778  0.4216
Dir_Speech  0.12769  0.3573   0.21
Game_Coding_Label (Intercept) 0.08960  0.2993
Dir_Speech  0.08216  0.2866  -0.17
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.3104      0.1648   1.883   0.0597 .
Dir_Speech  -0.4583      0.2042  -2.244   0.0248 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Dir_Speech -0.087

#####
#FRAMES OF REFERENCE
#####

#####
#Does frame of reference correlate with the presence of
  ↪ gesture
#####

Gesture.null <- glmer(Gesture ~ 1 + (1+FOR|Participant) +
  ↪ (1+FOR|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa")
  ↪ ))

Gesture.FOR <- glmer(Gesture ~ FOR + (1+FOR|Participant) +
  ↪ (1+FOR|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa")
  ↪ ))

```

```

anova(Gesture.null, Gesture.FOR)
Data: data
Models:
  Gesture.null: Gesture ~ 1 + (1 + FOR | Participant) + (1
    ↪ + FOR | Game_Coding_Label)
  Gesture.FOR: Gesture ~ FOR + (1 + FOR | Participant) + (1 +
    ↪ FOR | Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
Gesture.null 21 2242.2 2360.1 -1100.1  2200.2
Gesture.FOR  24 2242.3 2376.9 -1097.1  2194.3 5.973      3
    ↪      0.1129

#####
#Is frame of reference correlated with G_Speech... yes
#####

G_Speech.null <- glmer(G_Speech ~ 1 + (1|Participant) + (1+
  ↪ FOR|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

G_Speech.FOR <- glmer(G_Speech ~ FOR + (1|Participant) +
  ↪ (1+FOR|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

anova(G_Speech.null, G_Speech.FOR)
Data: data
Models:
  G_Speech.null: G_Speech ~ 1 + (1 | Participant) + (1 +
    ↪ FOR | Game_Coding_Label)
  G_Speech.FOR: G_Speech ~ FOR + (1 | Participant) + (1 + FOR
    ↪ | Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
G_Speech.null 12 2241.4 2308.7 -1108.7  2217.4
G_Speech.FOR  15 2226.3 2310.5 -1098.2  2196.3 21.112
    ↪      3  9.978e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

summary(G_Speech.FOR)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ FOR + (1 | Participant) + (1 + FOR |
  ↪ Game_Coding_Label)
Data: data

```

```
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2226.3   2310.5  -1098.1   2196.3     2006

Scaled residuals:
  Min      1Q  Median      3Q      Max
-7.1189 -0.6919  0.1957  0.7313  2.8551

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  0.26888  0.5185
Game_Coding_Label (Intercept)  0.14490  0.3807
FORGLO           0.17680  0.4205  -0.67
FORIN            0.80162  0.8953  -0.80  0.35
FORNone         0.08643  0.2940   0.39  0.30 -0.78
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.0461      0.2543   4.114 3.89e-05 ***
FORGLO       -0.7351      0.2718  -2.704 0.00684 **
FORIN        1.7088      0.4333   3.944 8.03e-05 ***
FORNone     -1.7580      0.2213  -7.946 1.93e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) FORGLO FORIN
FORGLO  -0.629
FORIN   -0.510  0.403
FORNone -0.460  0.595  0.045

lsmeans(G_Speech.FOR, pairwise~FOR, adjust ="tukey")
$lsmeans
FOR      lsmean      SE df  asymp.LCL  asymp.UCL
EGO     1.0460847  0.2542815  NA   0.5477022  1.5444673
GLO     0.3110227  0.2272571  NA  -0.1343931  0.7564384
IN       2.7548851  0.3741835  NA   2.0214988  3.4882713
None   -0.7119414  0.2487536  NA  -1.1994895 -0.2243932

Results are given on the logit (not the response) scale.
Confidence level used: 0.95

$contrasts
contrast      estimate      SE df  z.ratio p.value
EGO - GLO     0.7350621  0.2717927  NA    2.704  0.0345
```

```

EGO - IN    -1.7088003 0.4333196 NA   -3.944  0.0005
EGO - None  1.7580261 0.2212556 NA    7.946  <.0001
GLO - IN    -2.4438624 0.4084266 NA   -5.984  <.0001
GLO - None  1.0229641 0.2262943 NA    4.521  <.0001
IN - None   3.4668265 0.4776257 NA    7.258  <.0001

```

Results are given **on** the **log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

Tests are performed **on** the **log scale**

```
#####
```

```
#Is frame of reference correlated with P_Speech... yes
```

```
#####
```

```

P_Speech.null <- glmer(P_Speech ~ 1 + (1+FOR|Participant) +
  ↪ (1|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

```

```

P_Speech.FOR <- glmer(P_Speech ~ FOR + (1+FOR|Participant)
  ↪ + (1|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

```

```
anova(P_Speech.null, P_Speech.FOR)
```

```
Data: data
```

```
Models:
```

```

P_Speech.null: P_Speech ~ 1 + (1 + FOR | Participant) +
  ↪ (1 | Game_Coding_Label)

```

```

P_Speech.FOR: P_Speech ~ FOR + (1 + FOR | Participant) + (1
  ↪ | Game_Coding_Label)

```

```
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
P_Speech.null 12 1205.7 1273.0 -590.83 1181.7
```

```
P_Speech.FOR  15 1175.6 1259.7 -572.78 1145.6 36.092
```

```
↪      3 7.161e-08 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
↪ ' ' 1
```

```
summary(P_Speech.FOR)
```

```
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
```

```
Family: binomial (logit)
```

```
Formula: P_Speech ~ FOR + (1 + FOR | Participant) + (1 |
  ↪ Game_Coding_Label)
```

```
Data: data
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```

AIC      BIC    logLik deviance df.resid
1175.6   1259.7   -572.8   1145.6     2006

Scaled residuals:
  Min      1Q  Median      3Q      Max
-5.3369 -0.2447 -0.1762 -0.1212  8.2509

Random effects:
  Groups              Name          Variance Std.Dev. Corr
Participant          (Intercept)  0.4175   0.6462
FORGLO                0.1911   0.4371   0.60
FORIN                 1.1044   1.0509   0.27  0.59
FORNone              0.9544   0.9769  -0.47 -0.96 -0.78
Game_Coding_Label (Intercept)  0.4059   0.6371
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.8452      0.3967  -7.172 7.42e-13 ***
  FORGLO      3.6456      0.3360  10.850 < 2e-16 ***
  FORIN     -0.3646      0.6245  -0.584  0.559
FORNone      0.0198      0.4400   0.045  0.964
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) FORGLO FORIN
FORGLO  -0.487
FORIN   -0.256  0.484
FORNone -0.603  0.326  0.060

lsmeans(P_Speech.FOR, pairwise~FOR, adjust ="tukey")
$lsmeans
FOR      lsmean      SE df  asymp.LCL asymp.UCL
EGO  -2.8451837  0.3967346  NA  -3.62276928 -2.067598
GLO   0.8004408  0.3747462  NA   0.06595181  1.534930
IN   -3.2097458  0.6483311  NA  -4.48045147 -1.939040
None -2.8253790  0.3747953  NA  -3.55996423 -2.090794

Results are given on the logit (not the response) scale.
Confidence level used: 0.95

$contrasts
contrast      estimate      SE df  z.ratio p.value
EGO - GLO  -3.64562450  0.3360129  NA  -10.850 <.0001
EGO - IN    0.36456216  0.6244641  NA   0.584  0.9370
EGO - None -0.01980469  0.4399987  NA  -0.045  1.0000
GLO - IN    4.01018666  0.5475454  NA   7.324 <.0001

```



```
GLO - None  3.62581981 0.4583754 NA   7.910 <.0001
IN - None  -0.38436685 0.7421899 NA  -0.518  0.9548
```

Results are given **on** the **log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of

→ 4 estimates

Tests are performed **on** the **log scale**

```
#####
```

```
#Is frame of reference correlated with Dir in Gesture...
```

→ yes

```
#####
```

```
Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1+FOR|
  → Participant) + (1+FOR|Game_Coding_Label), data = data
  → , family = binomial, control = glmerControl(optimizer
  → = "bobyqa"))
```

```
Dir_Gesture.FOR <- glmer(Dir_Gesture ~ FOR + (1+FOR|
  → Participant) + (1+FOR|Game_Coding_Label), data = data
  → , family = binomial, control = glmerControl(optimizer
  → = "bobyqa"))
```

```
anova(Dir_Gesture.null, Dir_Gesture.FOR)
```

```
Data: data
```

```
Models:
```

```
Dir_Gesture.null: Dir_Gesture ~ 1 + (1 + FOR |
  → Participant) + (1 + FOR | Game_Coding_Label)
```

```
Dir_Gesture.FOR: Dir_Gesture ~ FOR + (1 + FOR | Participant
  → ) + (1 + FOR | Game_Coding_Label)
```

```
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
```

```
Dir_Gesture.null 21 1822.8 1940.6 -890.39 1780.8
```

```
Dir_Gesture.FOR  24 1817.2 1951.9 -884.62 1769.2 11.522
```

```
→      3    0.009214 **
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
→ ' ' 1
```

```
summary(Dir_Gesture.FOR)
```

```
Generalized linear mixed model fit by maximum likelihood (
  → Laplace Approximation) ['glmerMod']
```

```
Family: binomial (logit)
```

```
Formula: Dir_Gesture ~ FOR + (1 + FOR | Participant) + (1 +
  → FOR | Game_Coding_Label)
```

```
Data: data
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```
AIC      BIC    logLik deviance df.resid
1817.2    1951.9   -884.6   1769.2     1997
```

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-2.1948	-0.4701	-0.3177	-0.1248	5.2085

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	1.9769	1.4060	
FORGLO		0.3840	0.6197	-0.41
FORIN		0.3635	0.6029	-0.86
FORNone		0.7749	0.8803	-0.68
Game_Coding_Label	(Intercept)	1.0178	1.0089	
FORGLO		1.1351	1.0654	-0.65
FORIN		0.6726	0.8201	-0.91
FORNone		1.1300	1.0630	-0.97

Number of obs: 2021, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.6252	0.5517	-2.946	0.00322 **
FORGLO	-1.0346	0.6235	-1.659	0.09704 .
FORIN	-0.5185	0.4659	-1.113	0.26579
FORNone	0.5978	0.5085	1.176	0.23970

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 ↪ ' ' 1

Correlation of Fixed Effects:

	(Intr)	FORGLO	FORIN
FORGLO		-0.497	
FORIN		-0.732	0.670
FORNone		-0.833	0.618

lsmeans(Dir\_Gesture.FOR, pairwise~FOR, adjust ="tukey")

**\$lsmeans**

FOR	lsmean	SE	df	asyp.LCL	asyp.UCL
EGO	-1.625248	0.5517236	NA	-2.706606	-0.5438892
GLO	-2.659846	0.5927530	NA	-3.821621	-1.4980717
IN	-2.143698	0.3812863	NA	-2.891006	-1.3963907
None	-1.027402	0.3091701	NA	-1.633364	-0.4214393

Results are given **on** the logit (not the **response**) **scale**.

Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.ratio	p.value
EGO - GLO	1.0345985	0.6234947	NA	1.659	0.3454
EGO - IN	0.5184504	0.4658893	NA	1.113	0.6816
EGO - None	-0.5978461	0.5084822	NA	-1.176	0.6422
GLO - IN	-0.5161480	0.4654544	NA	-1.109	0.6840

```
GLO - None -1.6324445 0.5052525 NA -3.231 0.0068
IN - None -1.1162965 0.3790310 NA -2.945 0.0170
```

Results are given **on the log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of

→ 4 estimates

Tests are performed **on the log scale**

```
#####
```

```
#Is frame of reference correlated with M in Gesture... yes
```

```
#####
```

```
M_Gesture.null <- glmer(M_Gesture ~ 1 + (1+FOR|Participant)
  → + (1|Game_Coding_Label), data = data, family =
  → binomial, control = glmerControl(optimizer = "bobyqa"
  → ))
```

```
M_Gesture.FOR <- glmer(M_Gesture ~ FOR + (1+FOR|Participant)
  → ) + (1|Game_Coding_Label), data = data, family =
  → binomial, control = glmerControl(optimizer = "bobyqa"
  → ))
```

```
anova(M_Gesture.null, M_Gesture.FOR)
```

```
Data: data
```

```
Models:
```

```
  M_Gesture.null: M_Gesture ~ 1 + (1 + FOR | Participant) +
  → (1 | Game_Coding_Label)
```

```
  M_Gesture.FOR: M_Gesture ~ FOR + (1 + FOR | Participant) +
  → (1 | Game_Coding_Label)
```

```
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
M_Gesture.null 12 1803.2 1870.5 -889.59 1779.2
```

```
M_Gesture.FOR  15 1791.0 1875.1 -880.48 1761.0 18.224
```

```
  →      3  0.0003954 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
  → ' ' 1
```

```
summary(M_Gesture.FOR)
```

```
Generalized linear mixed model fit by maximum likelihood (
  → Laplace Approximation) ['glmerMod']
```

```
Family: binomial (logit)
```

```
Formula: M_Gesture ~ FOR + (1 + FOR | Participant) + (1 |
  → Game_Coding_Label)
```

```
Data: data
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```
AIC      BIC    logLik deviance df.resid
1791.0   1875.1  -880.5  1761.0    2006
```

```
Scaled residuals:
```

```

  Min      1Q  Median      3Q      Max
-2.9566 -0.4537 -0.2836 -0.1449  5.4456

```

Random **effects**:

```

  Groups              Name      Variance Std.Dev. Corr
Participant (Intercept) 1.9049   1.3802
FORGLO      0.3108    0.5575  -0.70
FORIN      0.5491    0.7410  -0.84  0.97
FORNone    0.1460    0.3821  -0.76  0.06  0.28
Game_Coding_Label (Intercept) 0.1650   0.4063
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

```

Fixed **effects**:

```

  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.92298    0.42848  -4.488 7.19e-06 ***
  FORGLO      -0.41981    0.31353  -1.339  0.181
FORIN        -0.09571    0.32005  -0.299  0.765
FORNone      1.09062    0.23577   4.626 3.73e-06 ***
  ---
  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

```

Correlation of Fixed Effects:

```

  (Intr) FORGLO FORIN
FORGLO  -0.529
FORIN   -0.662  0.683
FORNone -0.624  0.477  0.510

```

lsmeans(M\_Gesture.FOR, pairwise~FOR, adjust ="tukey")

**\$lsmeans**

```

FOR      lsmean      SE  df asymp.LCL  asymp.UCL
EGO -1.9229836 0.4284776 NA -2.762784 -1.0831828
GLO -2.3427978 0.3738337 NA -3.075498 -1.6100971
IN  -2.0186979 0.3232725 NA -2.652300 -1.3850955
None -0.8323596 0.3363700 NA -1.491633 -0.1730865

```

Results are given **on** the logit (not the **response**) **scale**.  
 Confidence level used: 0.95

**\$contrasts**

```

contrast      estimate      SE  df z.ratio p.value
EGO - GLO    0.41981418 0.3135271 NA  1.339  0.5380
EGO - IN     0.09571437 0.3200488 NA  0.299  0.9907
EGO - None  -1.09062396 0.2357706 NA -4.626 <.0001
GLO - IN    -0.32409981 0.2524136 NA -1.284  0.5731
GLO - None  -1.51043815 0.2886324 NA -5.233 <.0001
IN - None   -1.18633834 0.2845681 NA -4.169  0.0002

```

Results are given **on** the **log** (not the **response**) **scale**.

```

P value adjustment: tukey method for comparing a family of
  ↪ 4 estimates
Tests are performed on the log scale

#####
#Is frame of reference correlated with Dir in gesture
#####

Dir_Gesture.null <- glmer(Dir_Gesture~ 1 + (1+FOR|
  ↪ Participant) + (1+FOR|Game_Coding_Label), data = data
  ↪ , family = binomial, control = glmerControl(optimizer
  ↪ = "bobyqa"))

Dir_Gesture.FOR <- glmer(Dir_Gesture ~ FOR + (1+FOR|
  ↪ Participant) + (1+FOR|Game_Coding_Label), data = data
  ↪ , family = binomial, control = glmerControl(optimizer
  ↪ = "bobyqa"))

anova(Dir_Gesture.null, Dir_Gesture.FOR)
Data: data
Models:
  Dir_Gesture.null: Dir_Gesture ~ 1 + (1 + FOR |
    ↪ Participant) + (1 + FOR | Game_Coding_Label)
  Dir_Gesture.FOR: Dir_Gesture ~ FOR + (1 + FOR | Participant
    ↪ ) + (1 + FOR | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Gesture.null 21 1822.8 1940.6 -890.39   1780.8
Dir_Gesture.FOR  24 1817.2 1951.9 -884.62   1769.2 11.522
  ↪          3    0.009214

Dir_Gesture.null
Dir_Gesture.FOR  **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(Dir_Gesture.FOR)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula:
  Dir_Gesture ~ FOR + (1 + FOR | Participant) + (1 + FOR |
  ↪ Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
1817.2    1951.9   -884.6   1769.2     1997

```

```

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.1948 -0.4701 -0.3177 -0.1248  5.2085

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant          (Intercept) 1.9769   1.4060
FORGLO                0.3840   0.6197  -0.41
FORIN                 0.3635   0.6029  -0.86  0.68
FORNone              0.7749   0.8803  -0.68  0.53  0.43
Game_Coding_Label (Intercept) 1.0178   1.0089
FORGLO                1.1351   1.0654  -0.65
FORIN                 0.6726   0.8201  -0.91  0.91
FORNone              1.1300   1.0630  -0.97  0.80  0.98
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.6252      0.5517  -2.946  0.00322 **
  FORGLO      -1.0346      0.6235  -1.659  0.09704 .
FORIN        -0.5185      0.4659  -1.113  0.26579
FORNone       0.5978      0.5085   1.176  0.23970
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) FORGLO FORIN
FORGLO  -0.497
FORIN   -0.732  0.670
FORNone -0.833  0.618  0.701

lsmeans(Dir_Gesture.FOR, pairwise~FOR, adjust ="tukey")
$lsmeans
FOR      lsmean      SE  df  asymp.LCL  asymp.UCL
EGO  -1.625248  0.5517236 NA  -2.706606 -0.5438892
GLO  -2.659846  0.5927530 NA  -3.821621 -1.4980717
IN   -2.143698  0.3812863 NA  -2.891006 -1.3963907
None -1.027402  0.3091701 NA  -1.633364 -0.4214393

Results are given on the logit (not the response) scale.
Confidence level used: 0.95

$contrasts
contrast      estimate      SE  df  z.ratio  p.value
EGO - GLO     1.0345985  0.6234947 NA    1.659   0.3454
EGO - IN      0.5184504  0.4658893 NA    1.113   0.6816
EGO - None   -0.5978461  0.5084822 NA   -1.176   0.6422
GLO - IN     -0.5161480  0.4654544 NA   -1.109   0.6840

```

```
GLO - None -1.6324445 0.5052525 NA -3.231 0.0068
IN - None -1.1162965 0.3790310 NA -2.945 0.0170
```

Results are given **on** the **log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of

→ 4 estimates

Tests are performed **on** the **log scale**

```
#####
```

```
#Is frame of reference correlated with Dir in gesture
```

```
#####
```

```
Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+FOR|
  → Participant) + (1+FOR|Game_Coding_Label), data = data
  → , family = binomial, control = glmerControl(optimizer
  → = "bobyqa"))
```

```
Dir_Speech.FOR <- glmer(Dir_Speech ~ FOR + (1+FOR|
  → Participant) + (1+FOR|Game_Coding_Label), data = data
  → , family = binomial, control = glmerControl(optimizer
  → = "bobyqa"))
```

```
anova(Dir_Speech.null, Dir_Speech.FOR)
```

```
Data: data
```

```
Models:
```

```
Dir_Speech.null: Dir_Speech ~ 1 + (1 + FOR | Participant)
  → + (1 + FOR | Game_Coding_Label)
```

```
Dir_Speech.FOR: Dir_Speech ~ FOR + (1 + FOR | Participant)
  → + (1 + FOR | Game_Coding_Label)
```

```
Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
Dir_Speech.null 21 2229.3 2347.2 -1093.7 2187.3
```

```
Dir_Speech.FOR 24 2235.1 2369.8 -1093.5 2187.1 0.2436
```

```
→ 3 0.9703
```

```
summary(Dir_Speech.FOR)
```

```
Generalized linear mixed model fit by maximum likelihood (
```

```
→ Laplace Approximation) [glmerMod]
```

```
Family: binomial ( logit )
```

```
Formula:
```

```
Dir_Speech ~ FOR + (1 + FOR | Participant) + (1 + FOR |
  → Game_Coding_Label)
```

```
Data: data
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```
AIC BIC logLik deviance df.resid
```

```
2235.1 2369.8 -1093.6 2187.1 1997
```

```
Scaled residuals:
```

```
Min 1Q Median 3Q Max
-1.6845 -0.6284 -0.4411 0.7602 3.9583
```



Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.5003	0.7073	
FORGLO	0.5663 0.7526	-0.97		
FORIN	0.4807 0.6933	-0.60	0.68	
FORNone	1.0715 1.0351	-0.86	0.81	0.83
Game_Coding_Label	(Intercept)	0.5035	0.7095	
FORGLO	0.2727 0.5222	-0.30		
FORIN	0.4160 0.6450	-0.54	0.27	
FORNone	0.1467 0.3830	-0.42	0.36	0.98

Number of obs: 2021, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.49190	0.38496	-3.876	0.000106 ***
FORGLO	-0.18587	0.43312	-0.429	0.667811
FORIN	-0.01426	0.49657	-0.029	0.977092
FORNone	-0.11027	0.43720	-0.252	0.800875

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 ↪ ' ' 1

Correlation of Fixed Effects:

	(Intr)	FORGLO	FORIN
FORGLO	-0.668		
FORIN	-0.615	0.519	
FORNone	-0.728	0.638	0.722

lsmeans(Dir\_Speech.FOR, pairwise~FOR, adjust ="tukey")

**\$lsmeans**

FOR	lsmean	SE	df	asyp.LCL	asyp.UCL
EGO	-1.491902	0.3849565	NA	-2.246403	-0.7374013
GLO	-1.677776	0.3363445	NA	-2.336999	-1.0185531
IN	-1.506161	0.3996875	NA	-2.289534	-0.7227876
None	-1.602169	0.3069006	NA	-2.203684	-1.0006552

Results are given **on** the logit (not the **response**) **scale**.

Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.ratio	p.value
EGO - GLO	0.18587413	0.4331171	NA	0.429	0.9735
EGO - IN	0.01425855	0.4965662	NA	0.029	1.0000
EGO - None	0.11026723	0.4371954	NA	0.252	0.9944
GLO - IN	-0.17161558	0.4591172	NA	-0.374	0.9822
GLO - None	-0.07560690	0.3705217	NA	-0.204	0.9970
IN - None	0.09600868	0.3525056	NA	0.272	0.9929



```

Results are given on the log (not the response) scale.
P value adjustment: tukey method for comparing a family of
  ↪ 4 estimates
Tests are performed on the log scale
#####
#PERSPECTIVE
#####
#####
# Is Perspective correlated with G_Speech
#####
G_Speech.null <- glmer(G_Speech ~ 1 + (1|Participant) + (1|
  ↪ Game_Coding_Label), data = data_pers, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

G_Speech.perspective <- glmer(G_Speech ~ perspective + (1|
  ↪ Participant) + (1|Game_Coding_Label), data = data_
  ↪ pers, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.perspective)
Data: data_pers
Models:
  G_Speech.null: G_Speech ~ 1 + (1 | Participant) + (1 |
    ↪ Game_Coding_Label)
  G_Speech.perspective: G_Speech ~ perspective + (1 |
    ↪ Participant) + (1 | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Speech.null          3 2733.3 2750.2 -1363.7   2727.3
G_Speech.perspective  7 2639.2 2678.4 -1312.6   2625.2
  ↪ 102.14      4 < 2.2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(G_Speech.perspective)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ perspective + (1 | Participant) + (1 |
  ↪ Game_Coding_Label)
Data: data_pers
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2639.2   2678.4  -1312.6   2625.2     2002

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.9773 -0.9183  0.5991  0.8334  2.0680

```

Random **effects**:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.02788	0.1670
Game_Coding_Label	(Intercept)	0.11083	0.3329

Number of obs: 2009, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-1.1388	0.2723	-4.182	2.89e-05	***
perspectiveFP	0.8656	0.2525	3.428	0.000608	***
perspectivenone	1.5809	0.2701	5.853	4.83e-09	***
perspectiveS	1.9135	0.2561	7.470	8.01e-14	***
perspectiveSP	0.5739	0.3558	1.613	0.106734	

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	prspFP	prspct	prspcS
perspectvFP	-0.840			
perspectvnn	-0.833	0.874		
perspectivS	-0.799	0.847	0.817	
perspectvSP	-0.476	0.511	0.473	0.511

lsmeans(G\_Speech.perspective, pairwise~perspective, adjust  
 ↪ ="tukey")

**\$lsmeans**

	lsmean	SE	df	asyp.LCL	asyp.UCL
B	-1.1388344	0.2723189	NA	-1.6725696	-0.60509921
FP	-0.2732307	0.1497575	NA	-0.5667499	0.02028855
none	0.4420445	0.1567899	NA	0.1347419	0.74934718
S	0.7746378	0.1680561	NA	0.4452538	1.10402182
SP	-0.5649452	0.3293963	NA	-1.2105501	0.08065973

Results are given **on** the logit (not the **response**) **scale**.  
 Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.ratio	p.value
B - FP	-0.8656038	0.2525005	NA	-3.428	0.0055
B - none	-1.5808790	0.2700959	NA	-5.853	<.0001
B - S	-1.9134723	0.2561484	NA	-7.470	<.0001
B - SP	-0.5738892	0.3557797	NA	-1.613	0.4889
FP - none	-0.7152752	0.1320833	NA	-5.415	<.0001
FP - S	-1.0478685	0.1405887	NA	-7.453	<.0001
FP - SP	0.2917145	0.3139967	NA	0.929	0.8857
none - S	-0.3325933	0.1597037	NA	-2.083	0.2276

```

none - SP  1.0069897 0.3295988 NA  3.055  0.0191
S - SP    1.3395830 0.3147563 NA  4.256  0.0002

```

Results are given **on** the **log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of

→ 5 estimates

Tests are performed **on** the **log scale**

```

#####
#Is Dir_Speech correlated with perspective? NO
#####

```

```

Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1|Participant) +
  → (1+perspective|Game_Coding_Label), data = data_pers,
  → family = binomial, control = glmerControl(optimizer
  → = "bobyqa"))

```

```

Dir_Speech.perspective <- glmer(Dir_Speech ~ perspective +
  → (1|Participant) + (1+perspective|Game_Coding_Label),
  → data = data_pers, family = binomial, control =
  → glmerControl(optimizer = "bobyqa"))

```

```
anova(Dir_Speech.null, Dir_Speech.perspective)
```

```
Data: data_pers
```

```
Models:
```

```
Dir_Speech.null: Dir_Speech ~ 1 + (1 | Participant) + (1
  → + perspective | Game_Coding_Label)
```

```
Dir_Speech.perspective: Dir_Speech ~ perspective + (1 |
  → Participant) + (1 + perspective | Dir_Speech.
  → perspective: Game_Coding_Label)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
Dir_Speech.null			17 2228.8	2324.1	-1097.4			2194.8
Dir_Speech.perspective			21 2230.6	2348.3	-1094.3			2188.6
	→ 6.2086	4	0.1841					

```
summary(Dir_Speech.perspective)
```

```
Generalized linear mixed model fit by maximum likelihood (
  → Laplace Approximation) ['glmerMod']
```

```
Family: binomial ( logit )
```

```
Formula: Dir_Speech ~ perspective + (1 | Participant) + (1
  → + perspective | Game_Coding_Label)
```

```
Data: data_pers
```

```
Control: glmerControl(optimizer = "bobyqa")
```

AIC	BIC	logLik	deviance	df.resid
2230.6	2348.3	-1094.3	2188.6	1988

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-1.1239 -0.6409 -0.4457 0.8898 5.7805

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.1980	0.4450	
Game_Coding_Label	(Intercept)	1.0362	1.0180	
perspectiveFP	1.5794 1.2567	-0.83		
perspectivenone	2.8120 1.6769	-0.77 0.97		
perspectivesS	1.4516 1.2048	-0.78 0.98 0.93		
perspectiveSP	0.7495 0.8657	-0.02 -0.51 -0.62 -0.51		

Number of obs: 2009, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.8058	0.7068	-1.140 0.254
perspectiveFP	-0.6688	0.7355	-0.909 0.363
perspectivenone	-1.0965	0.8325	-1.317 0.188
perspectivesS	-0.8617	0.7429	-1.160 0.246
perspectiveSP	1.1461	0.8657	1.324 0.186

Correlation of Fixed Effects:

(Intr)	prspFP	prspct	prspcS
perspectvFP	-0.917		
perspectvnn	-0.860	0.931	
perspectivS	-0.893	0.941	0.894
perspectvSP	-0.497	0.378	0.291 0.379

lsmeans(Dir\_Speech.perspective, pairwise~perspective, adjust="tukey")

\$lsmeans

perspective	lsmean	SE	df	asyp.LCL	asyp.UCL
B	-0.8057790	0.7067883	NA	-2.191059	0.5795007
FP	-1.4745517	0.2952068	NA	-2.053146	-0.8959570
none	-1.9022435	0.4255263	NA	-2.736260	-1.0682272
S	-1.6674675	0.3378707	NA	-2.329682	-1.0052532
SP	0.3403443	0.8003212	NA	-1.228256	1.9089450

Results are given **on** the logit (not the **response**) **scale**.

Confidence level used: 0.95

\$contrasts

contrast	estimate	SE	df	z.ratio	p.value
B - FP	0.6687727	0.7355210	NA	0.909	0.8935
B - none	1.0964645	0.8325362	NA	1.317	0.6806
B - S	0.8616885	0.7429449	NA	1.160	0.7743
B - SP	-1.1461233	0.8656830	NA	-1.324	0.6762
FP - none	0.4276918	0.3054830	NA	1.400	0.6276
FP - S	0.1929158	0.2546180	NA	0.758	0.9426

```
FP - SP    -1.8148960 0.8995127 NA  -2.018  0.2574
none - S   -0.2347759 0.3732822 NA  -0.629  0.9704
none - SP  -2.2425878 1.0115188 NA  -2.217  0.1733
S - SP     -2.0078118 0.9023434 NA  -2.225  0.1703
```

Results are given **on** the **log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of

→ 5 estimates

Tests are performed **on** the **log scale**

```
#####
#Does pers have sig effect on M_Speech? NO
#####
```

```
M_Speech.null <- glmer(M_Speech ~ 1 + (1+perspective|
  → Participant) + (1|Game_Coding_Label), data = data_
  → pers, family = binomial, control = glmerControl(
  → optimizer = "bobyqa"))
```

```
M_Speech.perspective <- glmer(M_Speech ~ perspective + (1+
  → perspective|Participant) + (1|Game_Coding_Label),
  → data = data_pers, family = binomial, control =
  → glmerControl(optimizer = "bobyqa"))
```

```
anova(M_Speech.null, M_Speech.perspective)
```

```
Data: data_pers
```

```
Models:
```

```
  M_Speech.null: M_Speech ~ 1 + (1 + perspective |
  → Participant) + (1 | Game_Coding_Label)
```

```
M_Speech.perspective: M_Speech ~ perspective + (1 +
  → perspective | Participant) + (1 | M_Speech.
  → perspective: Game_Coding_Label)
```

```
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
M_Speech.null          17 2380.2 2475.4 -1173.1 2346.2
```

```
M_Speech.perspective  21 2384.4 2502.1 -1171.2 2342.4
```

```
  → 3.7756      4      0.4372
```

```
summary(M_Speech.perspective)
```

```
Generalized linear mixed model fit by maximum likelihood (
  → Laplace Approximation) ['glmerMod']
```

```
Family: binomial ( logit )
```

```
Formula: M_Speech ~ perspective + (1 + perspective |
  → Participant) + (1 | Game_Coding_Label)
```

```
Data: data_pers
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```
AIC      BIC    logLik deviance df.resid
2384.4    2502.1 -1171.2 2342.4    1988
```

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-1.1008	-0.6895	-0.4924	1.1221	3.5980

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.2808	0.5299	
perspectiveFP	0.1939	0.4403	-0.82	
perspectivenone	1.0754	1.0370	-1.00	0.79
perspectivesS	0.5988	0.7738	-0.76	0.26 0.79
perspectiveSP	0.2384	0.4883	-1.00	0.79 1.00 0.80
Game_Coding_Label	(Intercept)	0.4431	0.6657	

Number of obs: 2009, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.70709	0.52854	-3.230 0.00124 **
perspectiveFP	0.49872	0.48463	1.029 0.30345
perspectivenone	0.27585	0.55537	0.497 0.61940
perspectivesS	-0.04799	0.52264	-0.092 0.92683
perspectiveSP	0.45341	1.27675	0.355 0.72250

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 → ' ' 1

Correlation of Fixed Effects:

(Intr)	prspFP	prspct	prspcS
perspectvFP	-0.875		
perspectvnm	-0.856	0.886	
perspectivS	-0.827	0.839	0.873
perspectvSP	-0.306	0.324	0.318 0.310

lsmeans(M\_Speech.perspective, pairwise~perspective, adjust  
 → ="tukey")

**\$lsmeans**

perspective	lsmean	SE	df	asympt.LCL	asympt.UCL
B	-1.707093	0.5285444	NA	-2.743021	-0.6711648
FP	-1.208376	0.2566916	NA	-1.711483	-0.7052702
none	-1.431243	0.2917790	NA	-2.003119	-0.8593669
S	-1.755086	0.3089232	NA	-2.360564	-1.1496073
SP	-1.253685	1.2235432	NA	-3.651786	1.1444155

Results are given **on** the logit (not the **response**) **scale**.  
 Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.ratio	p.value
B - FP	-0.49871633	0.4846302	NA	-1.029	0.8420

```

B - none -0.27584960 0.5553707 NA -0.497 0.9877
B - S      0.04799283 0.5226377 NA  0.092 1.0000
B - SP    -0.45340773 1.2767539 NA -0.355 0.9966
FP - none  0.22286673 0.2579854 NA  0.864 0.9101
FP - S     0.54670916 0.2884839 NA  1.895 0.3199
FP - SP    0.04530860 1.2098210 NA  0.037 1.0000
none - S   0.32384243 0.2734412 NA  1.184 0.7604
none - SP -0.17755813 1.2197911 NA -0.146 0.9999
S - SP    -0.50140056 1.2204522 NA -0.411 0.9940

```

Results are given **on the log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of

→ 5 estimates

Tests are performed **on the log scale**

```

#####
data_persNoS <-rbind(data_persFP, data_persSP, data_persB)
#####

```

```

#####
#Is perspective correlated with M_Gesture? NO
#####

```

```

M_Gesture.null <- glmer(M_Gesture ~ 1 + (1+perspective|
  → Participant) + (1|Game_Coding_Label), data = data_
  → persNoS, family = binomial, control = glmerControl(
  → optimizer = "bobyqa"))

```

```

M_Gesture.perspective <- glmer(M_Gesture ~ perspective +
  → (1+perspective|Participant) + (1|Game_Coding_Label),
  → data = data_persNoS, family = binomial, control =
  → glmerControl(optimizer = "bobyqa"))

```

```

anova(M_Gesture.null, M_Gesture.perspective)

```

Data: **data\_persNoS**

Models:

```

M_Gesture.null: M_Gesture ~ 1 + (1 + perspective |
  → Participant) + (1 | Game_Coding_Label)

```

```

M_Gesture.perspective: M_Gesture ~ perspective + (1 +
  → perspective | Participant) + (1 | M_Gesture.
  → perspective: Game_Coding_Label)

```

```

Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
M_Gesture.null          8 965.82 1003.2 -474.91  949.82
M_Gesture.perspective 10 968.95 1015.6 -474.47  948.95
  → 0.8714      2    0.6468

```

```

summary(M_Gesture.perspective)

```

Generalized linear mixed **model** fit **by** maximum likelihood (

→ Laplace Approximation) ['glmerMod']

Family: **binomial** ( logit )

Formula: M\_Gesture ~ perspective + (1 + perspective |



```

↪ Participant) + (1 | Game_Coding_Label)
Data: data_persNoS
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
968.9    1015.6   -474.5   948.9     778

Scaled residuals:
  Min       1Q   Median       3Q      Max
-2.8483 -0.8316  0.4446  0.7283  2.1405

Random effects:
  Groups                Name                Variance Std.Dev. Corr
Participant             (Intercept)         0.7144   0.8452
perspectiveFP 1.9315     1.3898    -1.00
perspectiveSP 0.6426     0.8016    -1.00  1.00
Game_Coding_Label (Intercept)         0.5611   0.7491
Number of obs: 788, groups: Participant, 16; Game_Coding_
↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.8411    0.5557   1.514   0.130
perspectiveFP -0.6677    0.6072  -1.099   0.272
perspectiveSP -0.5119    0.8580  -0.597   0.551

Correlation of Fixed Effects:
  (Intr) prspFP
perspectvFP -0.864
perspectvSP -0.470  0.508

lsmeans(M_Gesture.perspective, pairwise~perspective, adjust
↪ ="tukey")
$lsmeans
perspective    lsmean      SE df  asymp.LCL asymp.UCL
B              0.8411040 0.5556938 NA  -0.2480359 1.9302440
FP              0.1734373 0.3073093 NA  -0.4288780 0.7757525
SP              0.3292394 0.7726436 NA  -1.1851142 1.8435930

Results are given on the logit (not the response) scale.
Confidence level used: 0.95

$contrasts
contrast      estimate      SE df  z.ratio p.value
B - FP        0.6676668 0.6072364 NA    1.100  0.5144
B - SP        0.5118647 0.8579788 NA    0.597  0.8219
FP - SP       -0.1558021 0.7584966 NA   -0.205  0.9770

Results are given on the log (not the response) scale.
P value adjustment: tukey method for comparing a family of

```



→ 3 estimates

Tests are performed **on the log scale**

```
#####
#Are speech framed gestures correlated with pers? YES. More
  → SF with B.
#####

speechFramed.null <- glmer(speechFramed ~ 1 + (1|
  → Participant) + (1|Game_Coding_Label), data = data_
  → pers, family = binomial, control = glmerControl(
  → optimizer = "bobyqa"))

speechFramed.perspective <- glmer(speechFramed ~
  → perspective + (1|Participant) + (1|Game_Coding_Label)
  → , data = data_pers, family = binomial, control =
  → glmerControl(optimizer = "bobyqa"))

anova(speechFramed.null, speechFramed.perspective)
Data: data_pers
Models:
  speechFramed.null: speechFramed ~ 1 + (1 | Participant) +
    → (1 | Game_Coding_Label)
  speechFramed.perspective: speechFramed ~ perspective + (1 |
    → Participant) + (1 | Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
speechFramed.null          3 973.66 990.47 -483.83  967.66
speechFramed.perspective  7 789.30 828.54 -387.65  775.30
  → 192.35      4 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  → ' ' 1

summary(speechFramed.perspective)
Generalized linear mixed model fit by maximum likelihood (
  → Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: speechFramed ~ perspective + (1 | Participant) +
  → (1 | Game_Coding_Label)
Data: data_pers
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
789.3    828.5   -387.7   775.3    2002

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.8498 -0.2005 -0.0688 -0.0338  6.8451
```

Random **effects**:

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	1.46577	1.2107
Game_Coding_Label	(Intercept)	0.07487	0.2736

Number of obs: 2009, groups: Participant, 16; Game\_Coding\_Label, 13

Fixed **effects**:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.5794	0.5161	-1.123	0.2616
perspectiveFP	-2.0055	0.3556	-5.639	1.71e-08 ***
perspectivenone	-5.4999	0.7777	-7.072	1.53e-12 ***
perspectiveS	-4.2850	0.4883	-8.775	< 2e-16 ***
perspectiveSP	-0.9560	0.3757	-2.544	0.0109 *

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 ↪ ' ' 1

Correlation of Fixed Effects:

	(Intr)	prspFP	prspct	prspcS
perspectvFP	-0.610			
perspectvnn	-0.317	0.405		
perspectivS	-0.423	0.621	0.278	
perspectvSP	-0.153	0.201	0.095	0.149

lsmeans(speechFramed.perspective, pairwise~perspective,  
 ↪ adjust ="tukey")

**\$lsmeans**

	lsmean	SE	df	asyp.LCL	asyp.UCL
B	-0.5793654	0.5160636	NA	-1.590831	0.4321006
FP	-2.5848911	0.4107432	NA	-3.389933	-1.7798492
none	-6.0793119	0.7854804	NA	-7.618825	-4.5397986
S	-4.8643467	0.5399821	NA	-5.922692	-3.8060013
SP	-1.5353449	0.5900018	NA	-2.691727	-0.3789625

Results are given **on** the logit (not the **response**) **scale**.  
 Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.ratio	p.value
B - FP	2.0055257	0.3556235	NA	5.639	<.0001
B - none	5.4999465	0.7776972	NA	7.072	<.0001
B - S	4.2849813	0.4883207	NA	8.775	<.0001
B - SP	0.9559795	0.3757302	NA	2.544	0.0811
FP - none	3.4944208	0.7120324	NA	4.908	<.0001
FP - S	2.2794556	0.3862854	NA	5.901	<.0001
FP - SP	-1.0495462	0.4625052	NA	-2.269	0.1549
none - S	-1.2149652	0.7951066	NA	-1.528	0.5441
none - SP	-4.5439670	0.8309724	NA	-5.468	<.0001

```
S - SP      -3.3290019 0.5700457 NA  -5.840  <.0001
```

Results are given **on the log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 5 estimates

Tests are performed **on the log scale**

```
#####  
#MOTION  
#####
```

```
#####  
#is motion correlated with perspective. Yes  
#####
```

```
motion.null <- glmer(motion ~ 1 + (1|Participant) + (1+  
  ↪ perspective|Game_Coding_Label), data = data_pers,  
  ↪ family = binomial, control = glmerControl(optimizer =  
  ↪ "bobyqa"))
```

```
motion.perspective <- glmer(motion ~ perspective + (1|  
  ↪ Participant) + (1+perspective|Game_Coding_Label),  
  ↪ data = data_pers, family = binomial, control =  
  ↪ glmerControl(optimizer = "bobyqa"))
```

```
anova(motion.null, motion.perspective)
```

```
Data: data_pers
```

```
Models:
```

```
  motion.null: motion ~ 1 + (1 | Participant) + (1 +  
    ↪ perspective | Game_Coding_Label)
```

```
motion.perspective: motion ~ perspective + (1 | Participant  
  ↪ ) + (1 + perspective | motion.perspective: Game_  
  ↪ Coding_Label)
```

```
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)  
motion.null          17 2389.2 2484.5 -1177.6    2355.2  
motion.perspective  21 2377.0 2494.7 -1167.5    2335.0 20.279
```

```
  ↪          4 0.0004399 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  
  ↪ ' ' 1
```

```
summary(motion.perspective)
```

```
Generalized linear mixed model fit by maximum likelihood (  
  ↪ Laplace Approximation) ['glmerMod']
```

```
Family: binomial ( logit )
```

```
Formula: motion ~ perspective + (1 | Participant) + (1 +  
  ↪ perspective | Game_Coding_Label)
```

```
Data: data_pers
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```

AIC      BIC    logLik deviance df.resid
2377.0   2494.7  -1167.5   2335.0    1988

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.0115 -0.6804 -0.4037  0.8120  5.9706

Random effects:
  Groups              Name                Variance Std.Dev. Corr
Participant          (Intercept)          0.5806   0.7620
Game_Coding_Label    (Intercept)          0.6367   0.7980
perspectiveFP        0.2534   0.5034   -0.88
perspectivenone      0.2379   0.4878   -0.04  0.50
perspectiveS         0.6607   0.8128   -0.59  0.90  0.83
perspectiveSP        8.1101   2.8478    0.07 -0.40 -0.79 -0.70
Number of obs: 2009, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.7119    0.5406   1.317  0.18788
perspectiveFP    -0.9517    0.4723  -2.015  0.04389 *
  perspectivene  -2.0342    0.4706  -4.322 1.54e-05 ***
  perspectiveS   -1.7272    0.5380  -3.210  0.00133 **
  perspectiveSP   0.1726    1.3476   0.128  0.89807
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) prspFP prspct prspcS
perspectvFP -0.875
perspectvnn -0.714  0.857
perspectivS -0.776  0.874  0.836
perspectvSP -0.245  0.223  0.098  0.029

#####
#Is motion correlated with gesture?
#####
gesture.null <- glmer(Gesture ~ 1 + (1+motion|Participant)
  ↪ + (1+motion|Game_Coding_Label), data = data, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa
  ↪ "))

gesture.motion <- glmer(Gesture ~ motion + (1+motion|
  ↪ Participant) + (1+motion|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

anova(gesture.null, gesture.motion)

```

```

Data: data
Models:
  gesture.null: Gesture ~ 1 + (1 + motion | Participant) +
    ↪ (1 + motion | Game_Coding_Label)
  gesture.motion: Gesture ~ motion + (1 + motion |
    ↪ Participant) + (1 + motion | gesture.motion: Game_
    ↪ Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
gesture.null    7 2230.4 2269.7 -1108.2  2216.4
gesture.motion  8 2219.5 2264.4 -1101.7  2203.5 12.912
    ↪      1  0.0003265 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪  ' ' 1

summary(gesture.motion)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Gesture ~ motion + (1 + motion | Participant) + (1
  ↪ + motion | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2219.5    2264.4 -1101.7  2203.5    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.8813 -0.5917 -0.4212  0.6643  3.1305

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  1.082648 1.04050
motion           0.504829 0.71051  1.00
Game_Coding_Label (Intercept) 0.046272 0.21511
motion           0.004247 0.06517 -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.7544      0.2855 -2.642  0.00824 **
motion       1.0150      0.2283  4.447  8.73e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪  ' ' 1

Correlation of Fixed Effects:
(Intr)

```

```

motion 0.645

#####
#Is motion tied to a reduction in G_Speech?
#####
G_Speech.null <- glmer(G_Speech ~ 1 + (1+motion|Participant
  ↪ ) + (1+motion|Game_Coding_Label), data = data, family
  ↪ = binomial, control = glmerControl(optimizer = "
  ↪ bobyqa"))

G_Speech.motion <- glmer(G_Speech ~ motion + (1+motion|
  ↪ Participant) + (1+motion|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.motion)
Data: data
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + motion | Participant)
  ↪ + (1 + motion | Game_Coding_Label)
  G_Speech.motion: G_Speech ~ motion + (1 + motion |
  ↪ Participant) + (1 + motion | G_Speech.motion: Game_
  ↪ Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Speech.null    7 2721.5 2760.7 -1353.7   2707.5
G_Speech.motion  8 2720.1 2765.0 -1352.0   2704.1 3.3829
  ↪    1    0.06587 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(G_Speech.motion)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ motion + (1 + motion | Participant) +
  ↪ (1 + motion | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2720.1   2765.0  -1352.0   2704.1     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.6403 -1.0250  0.6723  0.8464  1.8505

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept) 0.11840  0.3441

```

```

motion      0.22674  0.4762  0.19
Game_Coding_Label (Intercept) 0.10175  0.3190
motion      0.02795  0.1672  -0.79
Number of obs: 2021, groups:  Participant, 16; Game_Coding_
  ↳ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.3560      0.1599  2.226  0.0260 *
motion      -0.3591      0.1828 -1.964  0.0495 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

Correlation of Fixed Effects:
  (Intr)
motion -0.273

#####
#Is motion tied to FOR?
#####
motion.null <- glmer(motion ~ 1 + (1+FOR|Participant) + (1+
  ↳ FOR|Game_Coding_Label), data = data, family =
  ↳ binomial, control = glmerControl(optimizer = "bobyqa"
  ↳ ))

motion.FOR <- glmer(motion ~ FOR + (1+FOR|Participant) +
  ↳ (1+FOR|Game_Coding_Label), data = data, family =
  ↳ binomial, control = glmerControl(optimizer = "bobyqa"
  ↳ ))

anova(motion.null, motion.FOR)
Data: data
Models:
  motion.null: motion ~ 1 + (1 + FOR | Participant) + (1 +
    ↳ FOR | Game_Coding_Label)
  motion.FOR: motion ~ FOR + (1 + FOR | Participant) + (1 +
    ↳ FOR | Game_Coding_Label)
Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
motion.null 21 2419.5 2537.3 -1188.8 2377.5
motion.FOR  24 2412.8 2547.5 -1182.4 2364.8 12.702 3
  ↳ 0.005328 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

summary(motion.FOR)
Generalized linear mixed model fit by maximum likelihood (
  ↳ Laplace Approximation) ['glmerMod']

```

```

Family: binomial ( logit )
Formula: motion ~ FOR + (1 + FOR | Participant) + (1 + FOR
  ↪ | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2412.8   2547.5 -1182.4   2364.8     1997

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.8151 -0.6959 -0.4018  0.8248  6.3359

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant          (Intercept) 0.73239  0.8558
FORGLO                0.24055  0.4905  -0.86
FORIN                 0.23029  0.4799   0.03  0.23
FORNone               0.59149  0.7691  -0.36  0.02  0.32
Game_Coding_Label (Intercept) 0.44697  0.6686
FORGLO                0.07086  0.2662   0.38
FORIN                 0.24861  0.4986  -0.52  0.59
FORNone               0.09932  0.3151  -0.93 -0.01  0.80
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.6351      0.3573  -1.778 0.075484 .
FORGLO      -1.1743      0.3295  -3.564 0.000365 ***
  FORIN         0.1448      0.3418   0.424 0.671763
FORNone      0.3167      0.3131   1.011 0.311824
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) FORGLO FORIN
FORGLO  -0.425
FORIN   -0.418  0.440
FORNone -0.627  0.326  0.541

#####
# Complexity of speech?
#####
#####
# Complexity of speech? and Gesture
#####
Gesture.null <- glmer(Gesture ~ 1 + (1+Speech_AB|

```



```

  ↪ Participant) + (1+Speech_AB|Game_Coding_Label), data
  ↪ = data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Gesture.Speech_AB <- glmer(Gesture ~ Speech_AB + (1+Speech_
  ↪ AB|Participant) + (1+Speech_AB|Game_Coding_Label),
  ↪ data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(Gesture.null, Gesture.Speech_AB)
Data: data
Models:
  Gesture.null: Gesture ~ 1 + (1 + Speech_AB | Participant)
    ↪ + (1 + Speech_AB | Gesture.null: Game_Coding_
    ↪ Label)
  Gesture.Speech_AB: Gesture ~ Speech_AB + (1 + Speech_AB |
    ↪ Participant) + (1 + Speech_AB | Gesture.Speech_AB:
    ↪ Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Gesture.null      7 2287.8 2327.1 -1136.9  2273.8
Gesture.Speech_AB  8 2275.3 2320.2 -1129.6  2259.3 14.528
  ↪      1  0.0001381 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪  ' ' 1

summary(Gesture.Speech_AB)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial (logit)
Formula: Gesture ~ Speech_AB + (1 + Speech_AB | Participant
  ↪ ) + (1 + Speech_AB | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2275.3   2320.2 -1129.6  2259.3    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-4.5372 -0.5857 -0.4028  0.7282  3.3594

Random effects:
  Groups              Name              Variance Std.Dev. Corr
Participant          (Intercept)  2.115463  1.45446
Speech_AB            0.028957  0.17017  -0.94
Game_Coding_Label (Intercept)  0.045311  0.21286
Speech_AB            0.001637  0.04046  -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

```

Fixed **effects**:

```

  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.19003    0.40062   0.474   0.635
Speech_AB   -0.35440    0.07923  -4.473  7.7e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

```

Correlation of Fixed Effects:

```

(Intr)
Speech_AB -0.740

```

```

#####
# Complexity of speech and perspective
#####

```

```

Speech_AB.null <- lmer(Speech_AB ~ 1 + (1|Participant) +
  ↪ (1|Game_Coding_Label), data = data_pers, REML = FALSE
  ↪ )

```

```

Speech_AB.perspective <- lmer(Speech_AB ~ perspective + (1|
  ↪ Participant) + (1|Game_Coding_Label), data = data_
  ↪ pers, REML = FALSE)

```

```

anova(Speech_AB.null, Speech_AB.perspective)

```

```

Data: data_pers

```

```

Models:

```

```

  Speech_AB.null: Speech_AB ~ 1 + (1 | Participant) + (1 |
  ↪ Game_Coding_Label)

```

```

Speech_AB.perspective: Speech_AB ~ perspective + (1 |
  ↪ Participant) + (1 | Game_Coding_Label)

```

```

Df    AIC    BIC logLik deviance Chisq Chi Df
Speech_AB.null          4 5462.3 5484.7 -2727.2  5454.3
Speech_AB.perspective  8 5399.2 5444.0 -2691.6  5383.2
  ↪ 71.103      4

```

```

Pr(>Chisq)

```

```

Speech_AB.null

```

```

Speech_AB.perspective  1.328e-14 ***

```

```

---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

```

```

summary(Speech_AB.perspective)

```

```

Linear mixed model fit by maximum likelihood ['lmerMod']

```

```

Formula:

```

```

  Speech_AB ~ perspective + (1 | Participant) + (1 | Game_
  ↪ Coding_Label)

```

```

Data: data_pers

```

```

AIC      BIC    logLik deviance df.resid
5399.2   5444.0  -2691.6   5383.2     2001

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.3453 -0.7282 -0.0156  0.5489  5.5089

Random effects:
  Groups              Name          Variance Std.Dev.
Participant (Intercept) 0.02075  0.1441
Game_Coding_Label (Intercept) 0.03083  0.1756
Residual                                0.83888  0.9159
Number of obs: 2009, groups: Participant, 16; Game_Coding_
  → Label, 13

Fixed effects:
  Estimate Std. Error t value
(Intercept)      1.1174      0.1203    9.291
perspectiveFP    0.3221      0.1031    3.125
perspectivenone 0.5485      0.1046    5.246
perspectiveS     0.7220      0.1045    6.907
perspectiveSP    0.2563      0.1476    1.736

Correlation of Fixed Effects:
  (Intr) prspFP prspct prspcS
perspectvFP -0.753
perspectvnn -0.749  0.859
perspectivS -0.717  0.819  0.810
perspectvSP -0.388  0.445  0.442  0.445

lsmeans(Speech_AB.perspective, pairwise~perspective, adjust
  → ="tukey")
$lsmeans
perspective  lsmean      SE      df lower.CL upper.CL
B           1.117374 0.12370882 106.18 0.8721142 1.362634
FP          1.439433 0.08353907  24.84 1.2673257 1.611539
none       1.665877 0.08469551  25.33 1.4915595 1.840194
S          1.839376 0.08941690  31.28 1.6570766 2.021676
SP         1.373653 0.15291707 241.11 1.0724290 1.674877

Confidence level used: 0.95

$contrasts
contrast      estimate      SE      df t.ratio p.value
B - FP      -0.32205833 0.10419272 1578.29 -3.091  0.0173
B - none    -0.54850251 0.10597564 1341.03 -5.176 <.0001
B - S       -0.72200207 0.10553288 1687.54 -6.841 <.0001
B - SP      -0.25627871 0.14792521 2009.44 -1.732  0.4141
FP - none   -0.22644418 0.05590174 1228.47 -4.051  0.0005
FP - S      -0.39994374 0.06314349 1543.69 -6.334 <.0001

```

```
FP - SP      0.06577962 0.13861022 1753.14    0.475  0.9896
none - S     -0.17349956 0.06561011   968.28   -2.644  0.0634
none - SP    0.29222380 0.13943886 1675.40    2.096  0.2223
S - SP       0.46572336 0.13878856 1837.90    3.356  0.0072
```

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 5 estimates

```
#####
# AROUND
#####
#####
# does around predict gesture? NO
#####
```

```
gesture.null <- glmer(Gesture ~ 1 + (1+around|Participant)
    ↪ + (1+around|Game_Coding_Label), data = data, family =
    ↪ binomial, control = glmerControl(optimizer = "bobyqa
    ↪ "))
```

```
gesture.around <- glmer(Gesture ~ around + (1+around|
    ↪ Participant) + (1+around|Game_Coding_Label), data =
    ↪ data, family = binomial, control = glmerControl(
    ↪ optimizer = "bobyqa"))
```

```
anova(gesture.null, gesture.around)
```

```
Data: data
```

```
Models:
```

```
gesture.null: Gesture ~ 1 + (1 + around | Participant) +
    ↪ (1 + around | Game_Coding_Label)
gesture.around: Gesture ~ around + (1 + around |
    ↪ Participant) + (1 + around | gesture.around: Game_
    ↪ Coding_Label)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
gesture.null		7	2304.1	2343.3	-1145.0	2290.1	
gesture.around		8	2304.8	2349.7	-1144.4	2288.8	1.2139
		↪ 1		↪ 0.2706			

```
summary(gesture.around)
```

```
Generalized linear mixed model fit by maximum likelihood (
    ↪ Laplace Approximation) ['glmerMod']
```

```
Family: binomial ( logit )
```

```
Formula: Gesture ~ around + (1 + around | Participant) + (1
    ↪ + around | Game_Coding_Label)
```

```
Data: data
```

```
Control: glmerControl(optimizer = "bobyqa")
```

AIC	BIC	logLik	deviance	df.resid
2304.8	2349.7	-1144.4	2288.8	2013

```

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.3143 -0.5718 -0.4330  0.8532  3.0330

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  1.46420  1.2100
around           0.36864  0.6072  0.39
Game_Coding_Label (Intercept)  0.02364  0.1538
around           0.14577  0.3818  -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↳ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.4530      0.3211  -1.411    0.158
around      0.4053      0.3384   1.198    0.231

Correlation of Fixed Effects:
  (Intr)
around 0.015

#####
# does gesture predict around?
#####

around.null <- glmer(around ~ 1 + (1+Gesture|Participant) +
  ↳ (1+Gesture|Game_Coding_Label), data = data, family =
  ↳ binomial, control = glmerControl(optimizer = "bobyqa
  ↳ "))

around.gesture <- glmer(around ~ Gesture + (1+Gesture|
  ↳ Participant) + (1+Gesture|Game_Coding_Label), data =
  ↳ data, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))

anova(around.null, around.gesture)
Data: data
Models:
  around.null: around ~ 1 + (1 + Gesture | Participant) +
    ↳ (1 + Gesture | Game_Coding_Label)
  around.gesture: around ~ Gesture + (1 + Gesture |
    ↳ Participant) + (1 + Gesture | around.gesture: Game_
    ↳ Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
around.null      7 1090.9 1130.2 -538.44    1076.9
around.gesture  8 1092.8 1137.7 -538.41    1076.8 0.049
  ↳      1    0.8247

```

```
#####
# is ground correlated with gesture in environment of
#   ↪ around?
#####

around.G_Speech.null <- glmer(G_Speech ~ 1 + (1+Gesture|
  ↪ Participant) + (1+Gesture|Game_Coding_Label), data =
  ↪ data_around, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

around.G_Speech.gesture <- glmer(G_Speech ~ Gesture + (1+
  ↪ Gesture|Participant) + (1+Gesture|Game_Coding_Label),
  ↪ data = data_around, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(around.G_Speech.null, around.G_Speech.gesture)
Data: data_around
Models:
  around.G_Speech.null: G_Speech ~ 1 + (1 + Gesture |
    ↪ Participant) + (1 + Gesture | Game_Coding_Label)
  around.G_Speech.gesture: G_Speech ~ Gesture + (1 + Gesture
    ↪ | Participant) + (1 + Gesture | around.G_Speech.
    ↪ gesture: Game_Coding_Label)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
around.G_Speech.null      7 208.53 230.14 -97.263  194.53
around.G_Speech.gesture  8 209.17 233.87 -96.586  193.17
  ↪ 1.354      1    0.2446

summary(around.G_Speech.gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ Gesture + (1 + Gesture | Participant) +
  ↪ (1 + Gesture | Game_Coding_Label)
Data: data_around
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
209.2    233.9   -96.6    193.2    154

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.0124 -0.6391  0.4969  0.6191  1.8510

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  0.07086  0.2662
Gesture          0.96366  0.9817   1.00
Game_Coding_Label (Intercept)  0.01929  0.1389
```

```

Gesture      0.62658  0.7916  -1.00
Number of obs: 162, groups:  Participant, 15; Game_Coding_
  ↳ Label, 10

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.9883      0.2855   3.461 0.000538 ***
Gesture      -0.8492      0.6498  -1.307 0.191289
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Gesture -0.409

#####
# Does ground predict gesture in the presence of around
#####

around.gesture.null <- glmer(Gesture ~ 1 + (1+G_Speech |
  ↳ Participant) + (1+G_Speech |Game_Coding_Label), data
  ↳ = data_around, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))

around.gesture.G_Speech <- glmer(Gesture ~ G_Speech + (1+G_
  ↳ Speech|Participant) + (1+G_Speech|Game_Coding_Label),
  ↳ data = data_around, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))

anova(around.gesture.null, around.gesture.G_Speech)

summary(around.gesture.G_Speech)

#####
# is dir speech correlated with gesture in environment of
  ↳ around?
#####

around.Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+Gesture
  ↳ |Participant) + (1+Gesture|Game_Coding_Label), data =
  ↳ data_around, family = binomial, control =
  ↳ glmerControl(optimizer = "bobyqa"))

around.Dir_Speech.gesture <- glmer(Dir_Speech ~ Gesture +
  ↳ (1+Gesture|Participant) + (1+Gesture|Game_Coding_
  ↳ Label), data = data_around, family = binomial,
  ↳ control = glmerControl(optimizer = "bobyqa"))

```

```

anova(around.Dir_Speech.null, around.Dir_Speech.gesture)
Data: data_around
Models:
  around.Dir_Speech.null: Dir_Speech ~ 1 + (1 + Gesture |
    ↪ Participant) + (1 + Gesture | around.Dir_Speech.
    ↪ null: Game_Coding_Label)
  around.Dir_Speech.gesture: Dir_Speech ~ Gesture + (1 +
    ↪ Gesture | Participant) + (1 + Gesture | around.Dir_
    ↪ Speech.gesture: Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
around.Dir_Speech.null      7 169.25 190.86 -77.625  155.25
around.Dir_Speech.gesture  8 170.34 195.04 -77.171  154.34
    ↪ 0.9082      1    0.3406

summary(around.Dir_Speech.gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Speech ~ Gesture + (1 + Gesture | Participant)
  ↪ + (1 + Gesture | Game_Coding_Label)
Data: data_around
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
170.3    195.0   -77.2    154.3      154

Scaled residuals:
  Min      1Q  Median      3Q      Max
-0.5582 -0.5017 -0.4617 -0.3914  2.5551

Random effects:
  Groups          Name          Variance Std.Dev.  Corr
Participant      (Intercept) 9.039e-15 9.508e-08
Gesture          7.781e-15 8.821e-08 -1.00
Game_Coding_Label (Intercept) 9.289e-02 3.048e-01
Gesture          2.293e-01 4.789e-01 -1.00
Number of obs: 162, groups: Participant, 15; Game_Coding_
  ↪ Label, 10

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.7796      0.4737 -3.757 0.000172 ***
Gesture      0.5338      0.6732  0.793 0.427800
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Gesture -0.874

```



```
#####
# is dir speech correlated with G_Speech in environment of
# → around?
#####

around.Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+G_
  → Speech|Participant) + (1+G_Speech|Game_Coding_Label),
  → data = data_around, family = binomial, control =
  → glmerControl(optimizer = "bobyqa"))

around.Dir_Speech.G_Speech <- glmer(Dir_Speech ~ G_Speech +
  → (1+G_Speech|Participant) + (1+G_Speech|Game_Coding_
  → Label), data = data_around, family = binomial,
  → control = glmerControl(optimizer = "bobyqa"))

anova(around.Dir_Speech.null, around.Dir_Speech.G_Speech)
Data: data_around
Models:
  around.Dir_Speech.null: Dir_Speech ~ 1 + (1 + G_Speech |
    → Participant) + (1 + G_Speech |around.Dir_Speech.
    → null: Game_Coding_Label)
  around.Dir_Speech.G_Speech: Dir_Speech ~ G_Speech + (1 + G_
    → Speech | Participant) + (1 + G_Speech | around.Dir_
    → Speech.G_Speech: Game_Coding_Label)
Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
around.Dir_Speech.null      7 168.86 190.47 -77.430
  → 154.86
around.Dir_Speech.G_Speech  8 170.81 195.51 -77.404
  → 154.81 0.0517      1      0.8202

summary(around.Dir_Speech.G_Speech)
Generalized linear mixed model fit by maximum likelihood (
  → Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Speech ~ G_Speech + (1 + G_Speech |
  → Participant) + (1 + G_Speech | Game_Coding_Label
  → )
Data: data_around
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
170.8    195.5   -77.4    154.8    154

Scaled residuals:
  Min      1Q  Median      3Q      Max
-0.5290 -0.4753 -0.4563 -0.4320  2.3148

Random effects:
  Groups              Name              Variance Std.Dev.  Corr
```

```

Participant      (Intercept) 4.240e-13 6.511e-07
G_Speech        1.922e-12 1.386e-06 -0.99
Game_Coding_Label (Intercept) 5.983e-02 2.446e-01
G_Speech        1.719e-01 4.146e-01 -1.00
Number of obs: 162, groups:  Participant, 15; Game_Coding_
  ↪ Label, 10

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.4338      0.4369 -3.282 0.00103 **
  G_Speech   -0.1335      0.5658 -0.236 0.81348
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ '.' 1

Correlation of Fixed Effects:
  (Intr)
G_Speech -0.864

#####
# is dir gesture correlated with around
#####

Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1+around|
  ↪ Participant) + (1+around|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Dir_Gesture.around <- glmer(Dir_Gesture ~ around + (1+
  ↪ around|Participant) + (1+around|Game_Coding_Label),
  ↪ data = data, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(Dir_Gesture.null, Dir_Gesture.around)
Data: data
Models:
  Dir_Gesture.null: Dir_Gesture ~ 1 + (1 + around |
    ↪ Participant) + (1 + around | Dir_Gesture.null:
    ↪ Game_Coding_Label)
  Dir_Gesture.around: Dir_Gesture ~ around + (1 + around |
    ↪ Participant) + (1 + around | Dir_Gesture.around:
    ↪ Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Gesture.null      7 1844.5 1883.8 -915.26    1830.5
Dir_Gesture.around    8 1840.2 1885.1 -912.11    1824.2 6.3027
  ↪      1    0.01206 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ '.' 1

```

```

summary(Dir_Gesture.around)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Gesture ~ around + (1 + around | Participant)
  ↪ + (1 + around | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
1840.2   1885.1   -912.1   1824.2     2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.7447 -0.5025 -0.3172 -0.2371  3.8929

Random effects:
  Groups              Name      Variance Std.Dev. Corr
Participant (Intercept) 1.11277  1.0549
around      0.59751  0.7730  -0.02
Game_Coding_Label (Intercept) 0.11244  0.3353
around      0.01449  0.1204  -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.5324      0.3085  -4.967 6.82e-07 ***
around      1.1068      0.3661   3.023 0.0025 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
around -0.162

#####
# is dir speech correlated with around
#####

Dir_Speech.null <- glmer(Dir_Speech ~ 1 + (1+around|
  ↪ Participant) + (1+around|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Dir_Speech.around <- glmer(Dir_Speech ~ around + (1+around|
  ↪ Participant) + (1+around|Game_Coding_Label), data =
  ↪ data, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

```

```

anova(Dir_Speech.null,Dir_Speech.around)
Data: data
Models:
  Dir_Speech.null: Dir_Speech ~ 1 + (1 + around |
    ↪ Participant) + (1 + around | Game_Coding_Label)
  Dir_Speech.around: Dir_Speech ~ around + (1 + around |
    ↪ Participant) + (1 + around | Dir_Speech.around:
    ↪ Game_Coding_Label)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Dir_Speech.null    7 2221.2 2260.5 -1103.6  2207.2
Dir_Speech.around  8 2222.4 2267.3 -1103.2  2206.4 0.7258
    ↪      1    0.3943

summary(Dir_Speech.around)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Speech ~ around + (1 + around | Participant) +
  ↪ (1 + around | Game_Coding_Label)
Data: data
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2222.5    2267.3  -1103.2  2206.5    2013

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.1299 -0.6033 -0.4602  0.8851  4.0669

Random effects:
  Groups          Name          Variance Std.Dev. Corr
Participant      (Intercept)  0.18075  0.4252
around           0.26307  0.5129  -1.00
Game_Coding_Label (Intercept)  0.54419  0.7377
around           0.09881  0.3143  -1.00
Number of obs: 2021, groups: Participant, 16; Game_Coding_
  ↪ Label, 13

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.5310      0.2797 -5.473 4.43e-08 ***
around      -0.3919      0.4540 -0.863  0.388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
around -0.358

```

```
#####
# INTERACTION
#####
#####
# MOVE TYPES
#####
data_move <- subset(data, Game_Coding_Label != "R-N")
data_move <- subset(data_move, Game_Coding_Label != "
  ↳ response")
#####
# Move type on Gesture
#####

Gesture.null <- glmer(Gesture ~ 1 + (1|Task) + (1|Game_
  ↳ Coding_Level), data = data_move, family = binomial,
  ↳ control = glmerControl(optimizer = "bobyqa"))

Gesture.Game_Coding_Label <- glmer(Gesture ~ Game_Coding_
  ↳ Label + (1|Task) + (1|Game_Coding_Level), data = data
  ↳ _move, family = binomial, control = glmerControl(
  ↳ optimizer = "bobyqa"))

anova(Gesture.null, Gesture.Game_Coding_Label)
Data: data_move
Models:
  Gesture.null: Gesture ~ 1 + (1 | Task) + (1 | Game_Coding
    ↳ _Level)
  Gesture.Game_Coding_Label: Gesture ~ Game_Coding_Label + (1
    ↳ | Task) + (1 | Game_Coding_Level)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
Gesture.null                3 2455.4 2472.3 -1224.7   2449.4
Gesture.Game_Coding_Label 13 2404.1 2477.0 -1189.0   2378.1
  ↳ 71.392    10  2.386e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

summary(Gesture.Game_Coding_Label)
Generalized linear mixed model fit by maximum likelihood (
  ↳ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Gesture ~ Game_Coding_Label + (1 | Task) + (1 |
  ↳ Game_Coding_Level)
Data: data_move
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2404.1    2477.0  -1189.0   2378.1     2006
```

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-2.4660	-0.6141	-0.4286	0.7318	3.6459

Random **effects**:

Groups	Name	Variance	Std.Dev.
Game_Coding_Level	(Intercept)	0.7968	0.8926
Task	(Intercept)	1.1941	1.0927

Number of obs: 2019, groups: Game\_Coding\_Level, 1043; Task  
 ↪ , 8

Fixed **effects**:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)			-0.59007	0.50508	-1.168
	↪ 0.2427				
Game_Coding_LabelAlign			0.54350	0.36914	1.472
	↪ 0.1409				
Game_Coding_LabelCheck			-0.73227	0.34381	-2.130
	↪ 0.0332 *				
Game_Coding_LabelClarify			0.35454	0.34806	1.019
	↪ 0.3084				
Game_Coding_LabelExp			-0.22508	0.37707	-0.597
	↪ 0.5506				
Game_Coding_LabelInst			0.53828	0.33826	1.591
	↪ 0.1115				
Game_Coding_LabelQ-Wh			0.01961	0.64207	0.030
	↪ 0.9756				
Game_Coding_LabelQ-Y/N			0.08503	0.48482	0.175
	↪ 0.8608				
Game_Coding_LabelR-Wh			0.19962	0.41095	0.486
	↪ 0.6271				
Game_Coding_LabelR-Y			1.35672	0.59196	2.292
	↪ 0.0219 *				
Game_Coding_LabelReady			-1.18018	0.94547	-1.248
	↪ 0.2119				

---

Signif. **codes**: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 ↪ ' ' 1

Correlation of Fixed Effects:

(Intr)	G_C_LA	Gm_Cdng_LblCh	Gm_Cdng_LblCl	G_C_LE	G_C_LI	G_C_LQ-W	G_C_LQ-Y	G_C_LR-W
Gm_Cdng_LbA	-0.537							
Gm_Cdng_LblCh	-0.568	0.774						
Gm_Cdng_LblCl	-0.566	0.775	0.823					
Gm_Cdng_LbE	-0.525	0.715	0.765	0.755				
Gm_Cdng_LbI	-0.596	0.817	0.843	0.845				
	↪ 0.789							
Gm_Cdn_LQ-W	-0.315	0.424	0.450	0.442				

```

    ↪ 0.418 0.468
Gm_C_LQ-Y/N -0.413 0.564 0.595 0.588
    ↪ 0.553 0.627 0.337
Gm_Cdn_LR-W -0.489 0.655 0.694 0.692
    ↪ 0.645 0.721 0.388 0.509
Gm_Cdn_LR-Y -0.340 0.457 0.480 0.484
    ↪ 0.444 0.501 0.259 0.343 0.409
Gm_Cdng_LbR -0.208 0.283 0.298 0.297
    ↪ 0.279 0.315 0.163 0.215 0.253
G_C_LR-Y
Gm_Cdng_LbA
Gm_Cdng_LblCh
Gm_Cdng_LblCl
Gm_Cdng_LbE
Gm_Cdng_LbI
Gm_Cdn_LQ-W
Gm_C_LQ-Y/N
Gm_Cdn_LR-W
Gm_Cdn_LR-Y
Gm_Cdng_LbR 0.179

```

```

lsmeans(Gesture.Game_Coding_Label, pairwise~Game_Coding_
    ↪ Label, adjust="tukey")

```

```

$lsmeans
Game_Coding_Label      lsmean      SE df  asymp.LCL
    ↪ asymp.UCL
Ack                    -0.59007308 0.5050794 NA -1.5800104
    ↪ 0.39986426
Align                  -0.04657584 0.4371446 NA -0.9033635
    ↪ 0.81021188
Check                  -1.32234193 0.4197768 NA -2.1450894
    ↪ -0.49959443
Clarify                -0.23553566 0.4211917 NA -1.0610562
    ↪ 0.58998490
Exp                    -0.81515379 0.4440511 NA -1.6854779
    ↪ 0.05517032
Inst                   -0.05179045 0.4073736 NA -0.8502280
    ↪ 0.74664708
Q-Wh                   -0.57046075 0.6803921 NA -1.9040047
    ↪ 0.76308321
Q-Y/N                  -0.50503967 0.5366214 NA -1.5567983
    ↪ 0.54671894
R-Wh                   -0.39045651 0.4702077 NA -1.3120466
    ↪ 0.53113357
R-Y                    0.76664664 0.6341025 NA -0.4761714
    ↪ 2.00946468
Ready                  -1.77025233 0.9750122 NA -3.6812412
    ↪ 0.14073651

```

Results are given **on** the logit (not the **response**) **scale**.

Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.ratio	p.value
Ack - Align	-0.543497244	0.3691373	NA	-1.472	0.9290
Ack - Check	0.732268853	0.3438097	NA	2.130	0.5556
Ack - Clarify	-0.354537418	0.3480630	NA	-1.019	0.9951
Ack - Exp	0.225080711	0.3770704	NA	0.597	1.0000
Ack - Inst	-0.538282625	0.3382647	NA	-1.591	0.8864
Ack - Q-Wh	-0.019612327	0.6420678	NA	-0.031	1.0000
Ack - Q-Y/N	-0.085033406	0.4848164	NA	-0.175	1.0000
Ack - R-Wh	-0.199616571	0.4109496	NA	-0.486	1.0000
Ack - R-Y	-1.356719721	0.5919627	NA	-2.292	0.4394
Ack - Ready	1.180179247	0.9454713	NA	1.248	0.9769
Align - Check	1.275766097	0.2405891	NA	5.303	<.0001
Align - Clarify	0.188959826	0.2412303	NA	0.783	0.9995
Align - Exp	0.768577956	0.2818381	NA	2.727	0.1876
Align - Inst	0.005214619	0.2157955	NA	0.024	1.0000
Align - Q-Wh	0.523884918	0.5895101	NA	0.889	0.9984
Align - Q-Y/N	0.458463838	0.4117698	NA	1.113	0.9902
Align - R-Wh	0.343880674	0.3262760	NA	1.054	0.9936
Align - R-Y	-0.813222476	0.5355701	NA	-1.518	0.9141
Align - Ready	1.723676491	0.9123852	NA	1.889	0.7247
Check - Clarify	-1.086806271	0.2057714	NA	-5.282	<.0001
Check - Exp	-0.507188142	0.2491585	NA	-2.036	0.6237
Check - Inst	-1.270551478	0.1914392	NA	-6.637	<.0001
Check - Q-Wh	-0.751881180	0.5761782	NA	-1.305	0.9683
Check - Q-Y/N	-0.817302259	0.3935504	NA	-2.077	0.5941
Check - R-Wh	-0.931885424	0.3015969	NA	-3.090	0.0732
Check - R-Y	-2.088988574	0.5226933	NA	-3.997	0.0031
Check - Ready	0.447910394	0.9044663	NA	0.495	1.0000
Clarify - Exp	0.579618129	0.2551061	NA	2.272	0.4533
Clarify - Inst	-0.183745207	0.1913495	NA	-0.960	0.9970
Clarify - Q-Wh	0.334925092	0.5794647	NA	0.578	1.0000
Clarify - Q-Y/N	0.269504012	0.3971547	NA	0.679	0.9999
Clarify - R-Wh	0.154920848	0.3035732	NA	0.510	1.0000
Clarify - R-Y	-1.002182302	0.5214863	NA	-1.922	0.7030
Clarify - Ready	1.534716665	0.9052703	NA	1.695	0.8384
Exp - Inst	-0.763363337	0.2354635	NA	-3.242	0.0466
Exp - Q-Wh	-0.244693038	0.5931634	NA	-0.413	1.0000
Exp - Q-Y/N	-0.310114117	0.4185387	NA	-0.741	0.9997
Exp - R-Wh	-0.424697282	0.3336119	NA	-1.273	0.9734
Exp - R-Y	-1.581800432	0.5427416	NA	-2.914	0.1183
Exp - Ready	0.955098536	0.9150521	NA	1.044	0.9941
Inst - Q-Wh	0.518670299	0.5686230	NA	0.912	0.9981
Inst - Q-Y/N	0.453249220	0.3792551	NA	1.195	0.9832
Inst - R-Wh	0.338666055	0.2879995	NA	1.176	0.9851
Inst - R-Y	-0.818437095	0.5140855	NA	-1.592	0.8861
Inst - Ready	1.718461873	0.8983063	NA	1.913	0.7089
Q-Wh - Q-Y/N	-0.065421079	0.6616654	NA	-0.099	1.0000



<b>Q-Wh - R-Wh</b>	-0.180004244	0.6133984	NA	-0.293	1.0000
<b>Q-Wh - R-Y</b>	-1.337107394	0.7522038	NA	-1.778	0.7937
<b>Q-Wh - Ready</b>	1.199791574	1.0528878	NA	1.140	0.9883
<b>Q-Y/N - R-Wh</b>	-0.114583165	0.4486233	NA	-0.255	1.0000
<b>Q-Y/N - R-Y</b>	-1.271686315	0.6234836	NA	-2.040	0.6208
<b>Q-Y/N - Ready</b>	1.265212653	0.9654922	NA	1.310	0.9673
<b>R-Wh - R-Y</b>	-1.157103150	0.5658197	NA	-2.045	0.6170
<b>R-Wh - Ready</b>	1.379795818	0.9308481	NA	1.482	0.9260
<b>R-Y - Ready</b>	2.536898968	1.0217577	NA	2.483	0.3145

Results are given **on the log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of

→ 11 estimates

Tests are performed **on the log scale**

#####

*#Does move type have an effect on ground in speech*

#####

```
G_Speech.null <- glmer(G_Speech ~ 1 + (1|Task) + (1|Game_
  → Coding_Level), data = data_move, family = binomial,
  → control = glmerControl(optimizer = "bobyqa"))
```

```
G_Speech.Game_Coding_Label <- glmer(G_Speech ~ Game_Coding_
  → Label + (1|Task) + (1|Game_Coding_Level), data = data
  → _move, family = binomial, control = glmerControl(
  → optimizer = "bobyqa"))
```

```
anova(G_Speech.null, G_Speech.Game_Coding_Label)
```

Data: **data\_move**

Models:

```
G_Speech.null: G_Speech ~ 1 + (1 | Task) + (1 | Game_
  → Coding_Level)
```

```
G_Speech.Game_Coding_Label: G_Speech ~ Game_Coding_Label +
  → (1 | Task) + (1 | Game_Coding_Level)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
----	-----	-----	--------	----------	-------	-----	----	------------

G_Speech.null			3	2753.3	2770.1	-1373.6		
				→ 2747.3				

G_Speech.Game_Coding_Label	13	2743.9	2816.9	-1359.0				
		→ 2717.9	29.317	10	0.001107	**		

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1  
 → ‘ ’ 1

```
summary(G_Speech.Game_Coding_Label)
```

Generalized linear mixed **model** fit **by** maximum likelihood (  
 → Laplace Approximation) ['glmerMod']

Family: **binomial** ( logit )

Formula: G\_Speech ~ Game\_Coding\_Label + (1 | Task) + (1 |  
 → Game\_Coding\_Level)

```

Data: data_move
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
2743.9   2816.9 -1359.0   2717.9     2006

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.7487 -0.9542  0.6205  0.8705  1.5916

Random effects:
  Groups          Name          Variance Std.Dev.
Game_Coding_Level (Intercept) 0.2767   0.5260
Task               (Intercept) 0.1992   0.4463
Number of obs: 2019, groups: Game_Coding_Level, 1043; Task
  ↪ , 8

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)          -0.5946    0.3279  -1.814
  ↪ 0.06975 .
Game_Coding_LabelAlign  1.3375    0.3317   4.033 5.52e
  ↪ -05 ***
  Game_Coding_LabelCheck  0.8122    0.3021   2.689
  ↪ 0.00717 **
  Game_Coding_LabelClarify 0.6509    0.3078   2.114
  ↪ 0.03449 *
  Game_Coding_LabelExp    0.5743    0.3311   1.735
  ↪ 0.08278 .
Game_Coding_LabelInst   0.7143    0.2974   2.402
  ↪ 0.01631 *
  Game_Coding_LabelQ-Wh   1.2200    0.5719   2.133
  ↪ 0.03290 *
  Game_Coding_LabelQ-Y/N  1.3064    0.4326   3.020
  ↪ 0.00253 **
  Game_Coding_LabelR-Wh   0.9943    0.3594   2.766
  ↪ 0.00567 **
  Game_Coding_LabelR-Y    0.2480    0.5010   0.495
  ↪ 0.62061
Game_Coding_LabelReady  1.5955    0.7767   2.054
  ↪ 0.03995 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) G_C_LA Gm_Cdng_LblCh Gm_Cdng_LblCl G_C_LE G_C_LI G
  ↪ _C_LQ-W G_C_LQ-Y G_C_LR-W
Gm_Cdng_LbA      -0.729
Gm_Cdng_LblCh -0.800  0.794

```

```

Gm_Cdng_LbICl -0.784  0.778  0.854
Gm_Cdng_LbE   -0.735  0.726  0.794          0.779
Gm_Cdng_LbI   -0.823  0.811  0.882          0.865
  ↪ 0.813
Gm_Cdn_LQ-W   -0.434  0.423  0.459          0.450
  ↪ 0.426  0.476
Gm_C_LQ-Y/N   -0.569  0.560  0.608          0.598
  ↪ 0.563  0.632  0.334
Gm_Cdn_LR-W   -0.683  0.669  0.730          0.717
  ↪ 0.673  0.748  0.399  0.522
Gm_Cdn_LR-Y   -0.488  0.475  0.523          0.513
  ↪ 0.478  0.529  0.276  0.365  0.440
Gm_Cdng_LbR   -0.309  0.307  0.334          0.328
  ↪ 0.309  0.344  0.179  0.236  0.283
G_C_LR-Y
Gm_Cdng_LbA
Gm_Cdng_LbICl
Gm_Cdng_LbE
Gm_Cdng_LbI
Gm_Cdn_LQ-W
Gm_C_LQ-Y/N
Gm_Cdn_LR-W
Gm_Cdn_LR-Y
Gm_Cdng_LbR      0.203

lsmeans(G_Speech.Game_Coding_Label, pairwise~Game_Coding_
  ↪ Label, adjust="tukey")
#####
#Does move type have an effect on manner in gesture
#####

M_Gesture.null <- glmer(M_Gesture ~ 1 + (1|Task) + (1|Game_
  ↪ Coding_Level), data = data_move, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

M_Gesture.Game_Coding_Label <- glmer(M_Gesture ~ Game_
  ↪ Coding_Label + (1|Task) + (1|Game_Coding_Level), data
  ↪ = data_move, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(M_Gesture.null, M_Gesture.Game_Coding_Label)
Data: data_move
Models:
  M_Gesture.null: M_Gesture ~ 1 + (1 | Task) + (1 | Game_
    ↪ Coding_Level)
  M_Gesture.Game_Coding_Label: M_Gesture ~ Game_Coding_Label
    ↪ + (1 | Task) + (1 | Game_Coding_Level)
Df    AIC    BIC    logLik deviance Chisq Chi Df Pr(>Chisq)
M_Gesture.null                3 2053.6 2070.5 -1023.82

```

```

↪ 2047.6
M_Gesture.Game_Coding_Label 13 1977.4 2050.3 -975.71
↪ 1951.4 96.216 10 3.108e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
↪ ' ' 1

summary(M_Gesture.Game_Coding_Label)
Generalized linear mixed model fit by maximum likelihood (
↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: M_Gesture ~ Game_Coding_Label + (1 | Task) + (1 |
↪ Game_Coding_Level)
Data: data_move
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
1977.4   2050.4  -975.7   1951.4     2006

Scaled residuals:
  Min       1Q   Median       3Q      Max
-1.7883 -0.4809 -0.3281 -0.1357  4.6333

Random effects:
  Groups                Name      Variance Std.Dev.
Game_Coding_Level (Intercept) 1.262    1.123
Task                (Intercept) 1.143    1.069
Number of obs: 2019, groups: Game_Coding_Level, 1043; Task
↪ , 8

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)                -1.12085    0.53306  -2.103
↪ 0.03549 *
  Game_Coding_LabelAlign    -0.58990    0.43755  -1.348
↪ 0.17760
  Game_Coding_LabelCheck    -0.92086    0.39634  -2.323
↪ 0.02016 *
  Game_Coding_LabelClarify   0.03935    0.39921   0.099
↪ 0.92148
  Game_Coding_LabelExp      -1.23294    0.45782  -2.693
↪ 0.00708 **
  Game_Coding_LabelInst      0.53364    0.38731   1.378
↪ 0.16825
  Game_Coding_LabelQ-Wh     -1.44697    0.94288  -1.535
↪ 0.12487
  Game_Coding_LabelQ-Y/N    -1.54861    0.67547  -2.293
↪ 0.02187 *
  Game_Coding_LabelR-Wh     -0.98522    0.50803  -1.939
↪ 0.05247 .

```

```

Game_Coding_LabelR-Y      1.02137      0.62093      1.645
  ↳ 0.09999 .
Game_Coding_LabelReady   -1.55005      1.23950     -1.250
  ↳ 0.21110
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ '.' 1

Correlation of Fixed Effects:
  (Intr) G_C_LA Gm_Cdng_LblCh Gm_Cdng_LblCl G_C_LE G_C_LI G
  ↳ _C_LQ-W G_C_LQ-Y G_C_LR-W
Gm_Cdng_LbA      -0.546
Gm_Cdng_LblCh   -0.597  0.742
Gm_Cdng_LblCl   -0.606  0.732  0.806
Gm_Cdng_LbE     -0.524  0.645  0.715          0.700
Gm_Cdng_LbI     -0.651  0.766  0.822          0.831
  ↳ 0.727
Gm_Cdn_LQ-W     -0.265  0.314  0.346          0.339
  ↳ 0.306  0.365
Gm_C_LQ-Y/N     -0.355  0.446  0.488          0.474
  ↳ 0.433  0.503  0.219
Gm_Cdn_LR-W     -0.477  0.573  0.636          0.625
  ↳ 0.562  0.646  0.280  0.393
Gm_Cdn_LR-Y     -0.409  0.465  0.510          0.522
  ↳ 0.445  0.547  0.220  0.298  0.400
Gm_Cdng_LbR     -0.193  0.234  0.256          0.254
  ↳ 0.227  0.270  0.111  0.153  0.201
G_C_LR-Y
Gm_Cdng_LbA
Gm_Cdng_LblCh
Gm_Cdng_LblCl
Gm_Cdng_LbE
Gm_Cdng_LbI
Gm_Cdn_LQ-W
Gm_C_LQ-Y/N
Gm_Cdn_LR-W
Gm_Cdn_LR-Y
Gm_Cdng_LbR      0.167

lsmeans(M_Gesture.Game_Coding_Label, pairwise~Game_Coding_
  ↳ Label, adjust="tukey")
$lsmeans
Game_Coding_Label      lsmean      SE df asymp.LCL
  ↳ asymp.UCL
Ack                    -1.12085446  0.5330627 NA -2.165638
  ↳ -0.07607075
Align                  -1.71075012  0.4698694 NA -2.631677
  ↳ -0.78982295
Check                  -2.04171742  0.4346626 NA -2.893641
  ↳ -1.18979428

```

Clarify	-1.08150375	0.4307330	NA	-1.925725
↔	-0.23728261			
Exp	-2.35379823	0.4878787	NA	-3.310023
↔	-1.39757358			
Inst	-0.58720972	0.4064289	NA	-1.383796
↔	0.20937630			
Q-Wh	-2.56781948	0.9521515	NA	-4.434002
↔	-0.70163677			
Q-Y/N	-2.66946682	0.6961466	NA	-4.033889
↔	-1.30504447			
R-Wh	-2.10607242	0.5328192	NA	-3.150379
↔	-1.06176601			
R-Y	-0.09948927	0.6317107	NA	-1.337619
↔	1.13864089			
Ready	-2.67090644	1.2509887	NA	-5.122799
↔	-0.21901365			

Results are given **on** the logit (not the **response**) **scale**.  
 Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.ratio	p.value
Ack - Align	0.589895660	0.4375516	NA	1.348	0.9602
Ack - Check	0.920862964	0.3963405	NA	2.323	0.4176
Ack - Clarify	-0.039350705	0.3992052	NA	-0.099	1.0000
Ack - Exp	1.232943772	0.4578242	NA	2.693	0.2028
Ack - Inst	-0.533644743	0.3873054	NA	-1.378	0.9539
Ack - Q-Wh	1.446965017	0.9428756	NA	1.535	0.9084
Ack - Q-Y/N	1.548612358	0.6754688	NA	2.293	0.4389
Ack - R-Wh	0.985217959	0.5080288	NA	1.939	0.6912
Ack - R-Y	-1.021365187	0.6209278	NA	-1.645	0.8629
Ack - Ready	1.550051976	1.2395038	NA	1.251	0.9766
Align - Check	0.330967304	0.3022078	NA	1.095	0.9914
Align - Clarify	-0.629246366	0.3084212	NA	-2.040	0.6204
Align - Exp	0.643048112	0.3775855	NA	1.703	0.8345
Align - Inst	-1.123540403	0.2858885	NA	-3.930	0.0041
Align - Q-Wh	0.857069356	0.9062334	NA	0.946	0.9974
Align - Q-Y/N	0.958716698	0.6198188	NA	1.547	0.9039
Align - R-Wh	0.395322299	0.4413126	NA	0.896	0.9983
Align - R-Y	-1.611260848	0.5693652	NA	-2.830	0.1466
Align - Ready	0.960156316	1.2141119	NA	0.791	0.9994
Check - Clarify	-0.960213669	0.2476256	NA	-3.878	0.0050
Check - Exp	0.312080808	0.3273014	NA	0.953	0.9972
Check - Inst	-1.454507707	0.2341980	NA	-6.211	<.0001
Check - Q-Wh	0.526102053	0.8874083	NA	0.593	1.0000
Check - Q-Y/N	0.627749395	0.5935116	NA	1.058	0.9934
Check - R-Wh	0.064354996	0.3990638	NA	0.161	1.0000
Check - R-Y	-1.942228151	0.5401887	NA	-3.595	0.0144
Check - Ready	0.629189012	1.2006328	NA	0.524	1.0000
Clarify - Exp	1.272294477	0.3360660	NA	3.786	0.0071

Clarify - Inst	-0.494294038	0.2291641	NA	-2.157	0.5359
Clarify - Q-Wh	1.486315722	0.8906951	NA	1.669	0.8517
Clarify - Q-Y/N	1.587963064	0.6001363	NA	2.646	0.2251
Clarify - R-Wh	1.024568665	0.4046617	NA	2.532	0.2857
Clarify - R-Y	-0.982014482	0.5349234	NA	-1.836	0.7588
Clarify - Ready	1.589402682	1.2017829	NA	1.323	0.9652
Exp - Inst	-1.766588515	0.3189495	NA	-5.539	<.0001
Exp - Q-Wh	0.214021245	0.9133784	NA	0.234	1.0000
Exp - Q-Y/N	0.315668586	0.6307546	NA	0.500	1.0000
Exp - R-Wh	-0.247725813	0.4541519	NA	-0.545	1.0000
Exp - R-Y	-2.254308959	0.5848527	NA	-3.854	0.0055
Exp - Ready	0.317108204	1.2201357	NA	0.260	1.0000
Inst - Q-Wh	1.980609760	0.8786885	NA	2.254	0.4660
Inst - Q-Y/N	2.082257102	0.5857285	NA	3.555	0.0166
Inst - R-Wh	1.518862703	0.3922545	NA	3.872	0.0051
Inst - R-Y	-0.487720444	0.5220598	NA	-0.934	0.9976
Inst - Ready	2.083696719	1.1946506	NA	1.744	0.8125
Q-Wh - Q-Y/N	0.101647342	1.0324765	NA	0.098	1.0000
Q-Wh - R-Wh	-0.461747057	0.9375028	NA	-0.493	1.0000
Q-Wh - R-Y	-2.468330204	1.0083625	NA	-2.448	0.3359
Q-Wh - Ready	0.103086960	1.4716891	NA	0.070	1.0000
Q-Y/N - R-Wh	-0.563394399	0.6665991	NA	-0.845	0.9990
Q-Y/N - R-Y	-2.569977546	0.7691371	NA	-3.341	0.0340
Q-Y/N - Ready	0.001439618	1.3177419	NA	0.001	1.0000
R-Wh - R-Y	-2.006583147	0.6255568	NA	-3.208	0.0517
R-Wh - Ready	0.564834017	1.2416668	NA	0.455	1.0000
R-Y - Ready	2.571417164	1.2901237	NA	1.993	0.6539

Results are given **on the log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of

→ 11 estimates

Tests are performed **on the log scale**

```
#####
```

```
#Does move type have an effect on direction in gesture
```

```
#####
```

```
Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1|Task) + (1|
  → Game_Coding_Level), data = data_move, family =
  → binomial, control = glmerControl(optimizer = "bobyqa"
  → ))
```

```
Dir_Gesture.Game_Coding_Label <- glmer(Dir_Gesture ~ Game_
  → Coding_Label + (1|Task) + (1|Game_Coding_Level), data
  → = data_move, family = binomial, control =
  → glmerControl(optimizer = "bobyqa"))
```

```
anova(Dir_Gesture.null, Dir_Gesture.Game_Coding_Label)
```

```
Data: data_move
```



Models:

Dir\_Gesture.**null**: Dir\_Gesture ~ 1 + (1 | Task) + (1 |  
 ↪ Game\_Coding\_Level)

Dir\_Gesture.Game\_Coding\_Label: Dir\_Gesture ~ Game\_Coding\_  
 ↪ Label + (1 | Task) + (1 | Game\_Coding\_Level)

	Df	AIC	BIC	logLik
		↪ <b>deviance</b>	Chisq	Chi Df
		↪ Pr(>Chisq)		

Dir\_Gesture.**null** 3 1924.2 1941.1 -959.13  
 ↪ 1918.2

Dir\_Gesture.Game\_Coding\_Label 13 1877.8 1950.8 -925.91  
 ↪ 1851.8 66.441 10 2.147e-10 \*\*\*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
 ↪ ' ' 1

**summary**(Dir\_Gesture.Game\_Coding\_Label)

Generalized linear mixed **model** fit **by** maximum likelihood (  
 ↪ Laplace Approximation) ['glmerMod']

Family: **binomial** (logit)

Formula: Dir\_Gesture ~ Game\_Coding\_Label + (1 | Task) + (1  
 ↪ | Game\_Coding\_Level)

Data: **data\_move**

Control: glmerControl(optimizer = "bobyqa")

AIC	BIC	logLik	<b>deviance</b>	<b>df.resid</b>
1877.8	1950.7	-925.9	1851.8	2006

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-1.8025	-0.4520	-0.3257	-0.1899	3.9948

Random **effects**:

Groups	Name	Variance	Std.Dev.
Game_Coding_Level	(Intercept)	1.351	1.162
Task	(Intercept)	1.034	1.017

Number of obs: 2019, groups: Game\_Coding\_Level, 1043; Task  
 ↪ , 8

Fixed **effects**:

Estimate	Std. Error	z value	Pr(> z )
(Intercept)		-1.6950	0.5594 -3.030
↪ 0.00245	**		
Game_Coding_LabelAlign		-0.1959	0.4866 -0.402
↪ 0.68731			
Game_Coding_LabelCheck		-0.2762	0.4432 -0.623
↪ 0.53312			
Game_Coding_LabelClarify		0.3315	0.4483 0.740
↪ 0.45958			
Game_Coding_LabelExp		-0.6146	0.4984 -1.233



```

    ↪ 0.21753
Game_Coding_LabelInst      0.8361      0.4359      1.918
    ↪ 0.05510 .
Game_Coding_LabelQ-Wh     -0.9189      0.9514     -0.966
    ↪ 0.33414
Game_Coding_LabelQ-Y/N   -0.4667      0.6485     -0.720
    ↪ 0.47168
Game_Coding_LabelR-Wh    -0.5360      0.5493     -0.976
    ↪ 0.32920
Game_Coding_LabelR-Y      1.6920      0.6600      2.564
    ↪ 0.01036 *
  Game_Coding_LabelReady  -14.3276     49.3192     -0.290
    ↪ 0.77143
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) G_C_LA Gm_Cdng_LblCh Gm_Cdng_LblCl G_C_LE G_C_LI G
    ↪ _C_LQ-W G_C_LQ-Y G_C_LR-W
Gm_Cdng_LbA      -0.609
Gm_Cdng_LblCh   -0.669  0.770
Gm_Cdng_LblCl   -0.667  0.758  0.834
Gm_Cdng_LbE     -0.599  0.686  0.755      0.742
Gm_Cdng_LbI     -0.715  0.796  0.859      0.852
    ↪ 0.775
Gm_Cdn_LQ-W     -0.323  0.362  0.396      0.387
    ↪ 0.358  0.417
Gm_C_LQ-Y/N     -0.466  0.537  0.581      0.572
    ↪ 0.526  0.613  0.286
Gm_Cdn_LR-W     -0.542  0.615  0.677      0.667
    ↪ 0.611  0.690  0.326  0.477
Gm_Cdn_LR-Y     -0.479  0.513  0.565      0.567
    ↪ 0.504  0.594  0.267  0.389  0.451
Gm_Cdng_LbR      0.024 -0.033 -0.036      -0.036
    ↪ -0.035 -0.036 -0.012  -0.034  -0.031
G_C_LR-Y
Gm_Cdng_LbA
Gm_Cdng_LblCh
Gm_Cdng_LblCl
Gm_Cdng_LbE
Gm_Cdng_LbI
Gm_Cdn_LQ-W
Gm_C_LQ-Y/N
Gm_Cdn_LR-W
Gm_Cdn_LR-Y
Gm_Cdng_LbR      -0.037

lsmeans(Dir_Gesture.Game_Coding_Label, pairwise~Game_Coding
    ↪ _Label, adjust="tukey")

```

**\$lsmeans**

Game_Coding_Label	lsmean	SE	df	asympt.LCL
↪ asymp.UCL				
Ack	-1.695044342	0.5594429	NA	-2.791532
↪ -0.59855632				
Align	-1.890905258	0.4670741	NA	-2.806354
↪ -0.97545679				
Check	-1.971261688	0.4212404	NA	-2.796878
↪ -1.14564571				
Clarify	-1.363516402	0.4236074	NA	-2.193772
↪ -0.53326107				
Exp	-2.309612653	0.4765988	NA	-3.243729
↪ -1.37549618				
Inst	-0.858951969	0.3929396	NA	-1.629100
↪ -0.08880443				
Q-Wh	-2.613929070	0.9347737	NA	-4.446052
↪ -0.78180634				
Q-Y/N	-2.161793417	0.6288085	NA	-3.394235
↪ -0.92935141				
R-Wh	-2.231046956	0.5309382	NA	-3.271667
↪ -1.19042730				
R-Y	-0.003030652	0.6282270	NA	-1.234333
↪ 1.22827161				
Ready	-16.022605301	49.3355215	NA	-112.718451
↪ 80.67323995				

Results are given **on** the logit (not the **response**) **scale**.  
 Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.ratio	p.value
Ack - Align	0.19586092	0.4865963	NA	0.403	1.0000
Ack - Check	0.27621735	0.4431850	NA	0.623	0.9999
Ack - Clarify	-0.33152794	0.4482861	NA	-0.740	0.9997
Ack - Exp	0.61456831	0.4983873	NA	1.233	0.9788
Ack - Inst	-0.83609237	0.4358981	NA	-1.918	0.7055
Ack - Q-Wh	0.91888473	0.9514131	NA	0.966	0.9969
Ack - Q-Y/N	0.46674907	0.6484826	NA	0.720	0.9998
Ack - R-Wh	0.53600261	0.5493337	NA	0.976	0.9966
Ack - R-Y	-1.69201369	0.6599908	NA	-2.564	0.2680
Ack - Ready	14.32756096	49.3191627	NA	0.291	1.0000
Align - Check	0.08035643	0.3177015	NA	0.253	1.0000
Align - Clarify	-0.52738886	0.3269686	NA	-1.613	0.8773
Align - Exp	0.41870740	0.3901249	NA	1.073	0.9926
Align - Inst	-1.03195329	0.2986669	NA	-3.455	0.0234
Align - Q-Wh	0.72302381	0.8982254	NA	0.805	0.9993
Align - Q-Y/N	0.27088816	0.5644305	NA	0.480	1.0000
Align - R-Wh	0.34014170	0.4579191	NA	0.743	0.9997
Align - R-Y	-1.88787461	0.5856571	NA	-3.224	0.0493
Align - Ready	14.13170004	49.3377218	NA	0.286	1.0000

Check - Clarify	-0.60774529	0.2566841	NA	-2.368	0.3877
Check - Exp	0.33835097	0.3332564	NA	1.015	0.9953
Check - Inst	-1.11230972	0.2338853	NA	-4.756	0.0001
Check - Q-Wh	0.64266738	0.8761660	NA	0.733	0.9997
Check - Q-Y/N	0.19053173	0.5321772	NA	0.358	1.0000
Check - R-Wh	0.25978527	0.4106568	NA	0.633	0.9999
Check - R-Y	-1.96823104	0.5491965	NA	-3.584	0.0150
Check - Ready	14.05134361	49.3371419	NA	0.285	1.0000
Clarify - Exp	0.94609625	0.3430245	NA	2.758	0.1745
Clarify - Inst	-0.50456443	0.2410379	NA	-2.093	0.5821
Clarify - Q-Wh	1.25041267	0.8809175	NA	1.419	0.9439
Clarify - Q-Y/N	0.79827701	0.5376986	NA	1.485	0.9253
Clarify - R-Wh	0.86753055	0.4175993	NA	2.077	0.5936
Clarify - R-Y	-1.36048575	0.5486688	NA	-2.480	0.3164
Clarify - Ready	14.65908890	49.3371266	NA	0.297	1.0000
Exp - Inst	-1.45066068	0.3187968	NA	-4.550	0.0003
Exp - Q-Wh	0.30431642	0.9024769	NA	0.337	1.0000
Exp - Q-Y/N	-0.14781924	0.5736747	NA	-0.258	1.0000
Exp - R-Wh	-0.07856570	0.4641515	NA	-0.169	1.0000
Exp - R-Y	-2.30658200	0.5938933	NA	-3.884	0.0049
Exp - Ready	13.71299265	49.3389834	NA	0.278	1.0000
Inst - Q-Wh	1.75497710	0.8658585	NA	2.027	0.6300
Inst - Q-Y/N	1.30284145	0.5135577	NA	2.537	0.2829
Inst - R-Wh	1.37209499	0.4017053	NA	3.416	0.0267
Inst - R-Y	-0.85592132	0.5329290	NA	-1.606	0.8802
Inst - Ready	15.16365333	49.3366081	NA	0.307	1.0000
Q-Wh - Q-Y/N	-0.45213565	0.9863330	NA	-0.458	1.0000
Q-Wh - R-Wh	-0.38288211	0.9308089	NA	-0.411	1.0000
Q-Wh - R-Y	-2.61089842	1.0028548	NA	-2.603	0.2467
Q-Wh - Ready	13.40867623	49.3394491	NA	0.272	1.0000
Q-Y/N - R-Wh	0.06925354	0.6185180	NA	0.112	1.0000
Q-Y/N - R-Y	-2.15876276	0.7234709	NA	-2.984	0.0984
Q-Y/N - Ready	13.86081188	49.3453283	NA	0.281	1.0000
R-Wh - R-Y	-2.22801630	0.6404557	NA	-3.479	0.0216
R-Wh - Ready	13.79155835	49.3394990	NA	0.280	1.0000
R-Y - Ready	16.01957465	49.3478678	NA	0.325	1.0000

Results are given **on the log** (not the **response**) **scale**.

P value adjustment: tukey method **for** comparing a **family** of  
 → 11 estimates

Tests are performed **on the log scale**

```
#####
#Does M gesture and dir gesture still negatively effect G_
# → speech in the environmnet of Explain moves.
#####
```

```
data_Expain <- subset(data, Game_Coding_Label == "Exp")
```

```
G_Speech.null <- glmer(G_Speech ~ 1 + (1 + M_Gesture |
```

```

  ↪ Participant), data = data_Explain, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

G_Speech.M_Gesture <- glmer(G_Speech ~ M_Gesture + (1 + M_
  ↪ Gesture |Participant), data = data_Explain, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

anova(G_Speech.null, G_Speech.M_Gesture)
Data: data_Explain
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + M_Gesture |
    ↪ Participant)
  G_Speech.M_Gesture: G_Speech ~ M_Gesture + (1 + M_Gesture |
    ↪ Participant)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Speech.null      4 221.57 233.92 -106.79  213.57
G_Speech.M_Gesture  5 209.28 224.72  -99.64  199.28 14.294
  ↪      1  0.0001564 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(G_Speech.M_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ M_Gesture + (1 + M_Gesture |
  ↪ Participant)
Data: data_Explain
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
209.3    224.7   -99.6    199.3      157

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.1326 -1.0915 -0.2053  0.8867  4.6626

Random effects:
  Groups      Name      Variance Std.Dev. Corr
Participant (Intercept) 0.0166912 0.12919
M_Gesture 0.0006904 0.02627 1.00
Number of obs: 162, groups: Participant, 15

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.2225 0.1818 1.224 0.22094
M_Gesture -3.3324 1.0983 -3.034 0.00241 **
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ' ' 1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
M_Gesture -0.173

#####
G_Speech.null <- glmer(G_Speech ~ 1 + (1|Participant), data
  ↪ = data_Explain, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

G_Speech.Dir_Gesture <- glmer(G_Speech ~ Dir_Gesture + (1|
  ↪ Participant), data = data_Explain, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.Dir_Gesture)
Data: data_Explain
Models:
  G_Speech.null: G_Speech ~ 1 + (1 | Participant)
  G_Speech.Dir_Gesture: G_Speech ~ Dir_Gesture + (1 |
  ↪ Participant)
Df    AIC    BIC   logLik deviance Chisq Chi Df Pr(>Chisq)
G_Speech.null          2 228.36 234.53 -112.179  224.36
G_Speech.Dir_Gesture  3 204.63 213.89  -99.314  198.63
  ↪ 25.729      1 3.928e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ' ' 1
  ↪ ' ' 1

summary(G_Speech.Dir_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ Dir_Gesture + (1 | Participant)
Data: data_Explain
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC   logLik deviance df.resid
204.6    213.9   -99.3   198.6     159

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.2849 -0.9856 -0.2024  0.8493  4.6436

Random effects:
  Groups      Name      Variance Std.Dev.
Participant (Intercept) 0.1207  0.3475
Number of obs: 162, groups: Participant, 15

```

```

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.2647      0.2185  1.212  0.22569
Dir_Gesture -3.5015      1.0780 -3.248  0.00116 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Dir_Gesture -0.214

#####
#M and Dir in Gesture affect G_Speech environemnt of Inst
#####
data_Inst <- subset(data, Game_Coding_Label == "Inst")

G_Speech.null <- glmer(G_Speech ~ 1 + (1|Participant), data
  ↪ = data_Inst, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

G_Speech.M_Gesture <- glmer(G_Speech ~ M_Gesture + (1|
  ↪ Participant), data = data_Inst, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.M_Gesture)
Data: data_Inst
Models:
  G_Speech.null: G_Speech ~ 1 + (1 | Participant)
  G_Speech.M_Gesture: G_Speech ~ M_Gesture + (1 | Participant
  ↪ )

              Df    AIC    BIC logLik deviance
              ↪ Chisq Chi Df Pr(>Chisq)
G_Speech.null      2 887.99 897.01 -442.00    883.99
G_Speech.M_Gesture 3 833.02 846.54 -413.51    827.02 56.977
  ↪          1 4.41e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                  ↪ ' ' 1

summary(G_Speech.M_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ M_Gesture + (1 | Participant)
Data: data_Inst
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
833.0    846.5   -413.5    827.0      667

```

```

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.6253 -0.6589  0.6153  0.7223  1.9175

Random effects:
  Groups      Name      Variance Std.Dev.
Participant (Intercept) 0.1046  0.3234
Number of obs: 670, groups: Participant, 9

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.5738      0.1657   3.464 0.000533 ***
M_Gesture   -1.4849      0.2002  -7.416 1.2e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
M_Gesture -0.422

##
G_Speech.null <- glmer(G_Speech ~ 1 + (1+Dir_Gesture|
  ↪ Participant), data = data_Inst, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

G_Speech.Dir_Gesture <- glmer(G_Speech ~ Dir_Gesture + (1+
  ↪ Dir_Gesture|Participant), data = data_Inst, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

anova(G_Speech.null, G_Speech.Dir_Gesture)
Data: data_Inst
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + Dir_Gesture |
  ↪ Participant)
  G_Speech.Dir_Gesture: G_Speech ~ Dir_Gesture + (1 + Dir_
  ↪ Gesture | Participant)
              Df    AIC    BIC logLik deviance
              ↪ Chisq Chi Df Pr(>Chisq)
G_Speech.null      4 865.90 883.93 -428.95  857.90
G_Speech.Dir_Gesture 5 859.82 882.36 -424.91  849.82
  ↪ 8.0766      1 0.004484 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

summary(G_Speech.Dir_Gesture)

```



```

Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ Dir_Gesture + (1 + Dir_Gesture |
  ↪ Participant)
Data: data_Inst
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
859.8    882.4   -424.9    849.8      665

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.5530 -1.0302  0.6439  0.8123  2.2125

Random effects:
  Groups      Name      Variance Std.Dev. Corr
Participant (Intercept) 0.1322  0.3636
Dir_Gesture 0.2050  0.4528  1.00
Number of obs: 670, groups: Participant, 9

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.4168    0.1676  2.486  0.0129 *
Dir_Gesture -1.0353    0.2602 -3.979 6.92e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ '.' 1

Correlation of Fixed Effects:
  (Intr)
Dir_Gesture 0.267

#####
#M and Dir in Gesture affect G_Speech environemnt of Checks
#####
data_check <- subset(data, Game_Coding_Label == "Check")

G_Speech.null <- glmer(G_Speech ~ 1 + (1 + M_Gesture |
  ↪ Participant), data = data_check, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

G_Speech.M_Gesture <- glmer(G_Speech ~ M_Gesture + (1 + M_
  ↪ Gesture |Participant), data = data_check, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

anova(G_Speech.null, G_Speech.M_Gesture)
Data: data_check

```



```

Models:
  G_Speech.null: G_Speech ~ 1 + (1 + M_Gesture |
    ↪ Participant)
  G_Speech.M_Gesture: G_Speech ~ M_Gesture + (1 + M_Gesture |
    ↪ Participant)
Df      AIC      BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
G_Speech.null      4 553.65 569.68 -272.83  545.65
G_Speech.M_Gesture 5 543.77 563.80 -266.89  533.77 11.88
    ↪      1 0.0005673 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪  ' ' 1

summary(G_Speech.M_Gesture)
Generalized linear mixed model fit by maximum likelihood (
    ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ M_Gesture + (1 + M_Gesture |
    ↪ Participant)
Data: data_check
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
543.8    563.8  -266.9   533.8      401

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.2157 -1.2157  0.8226  0.8226  1.6832

Random effects:
  Groups      Name      Variance Std.Dev.  Corr
Participant (Intercept) 0.000e+00 0.000e+00
M_Gesture      1.036e-14 1.018e-07  NaN
Number of obs: 406, groups: Participant, 12

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.3907      0.1110  3.518 0.000434 ***
M_Gesture    -1.4321      0.2958 -4.842 1.29e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪  ' ' 1

Correlation of Fixed Effects:
  (Intr)
M_Gesture -0.375

##
G_Speech.null <- glmer(G_Speech ~ 1 + (1 + Dir_Gesture |
    ↪ Participant), data = data_check, family = binomial,

```

```

↪ control = glmerControl(optimizer = "bobyqa"))

G_Speech.Dir_Gesture <- glmer(G_Speech ~ Dir_Gesture + (1 +
↪ Dir_Gesture | Participant), data = data_check, family
↪ = binomial, control = glmerControl(optimizer = "
↪ bobyqa"))

anova(G_Speech.null, G_Speech.Dir_Gesture)
Data: data_check
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + Dir_Gesture |
↪ Participant)
  G_Speech.Dir_Gesture: G_Speech ~ Dir_Gesture + (1 + Dir_
↪ Gesture | Participant)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Speech.null          4 560.46 576.49 -276.23   552.46
G_Speech.Dir_Gesture  5 554.37 574.41 -272.19   544.37
↪ 8.0882      1  0.004455 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
↪ ' ' 1

summary(G_Speech.Dir_Gesture)
Generalized linear mixed model fit by maximum likelihood (
↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ Dir_Gesture + (1 + Dir_Gesture |
↪ Participant)
Data: data_check
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
554.4    574.4   -272.2   544.4     401

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.1964 -1.1651  0.8359  0.8548  1.5004

Random effects:
  Groups      Name      Variance Std.Dev. Corr
Participant (Intercept) 0.006648 0.08154
Dir_Gesture 0.016852 0.12982 -1.00
Number of obs: 406, groups: Participant, 12

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.3136      0.1536  2.041 0.041238 *
Dir_Gesture -1.0984      0.3198 -3.434 0.000594 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

```

```

    ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Dir_Gesture -0.529

#####
#M and Dir in Gesture affect G_Speech environemnt of
  ↪ Clarify
#####
data_Clarify <- subset(data, Game_Coding_Label == "Clarify"
  ↪ )

G_Speech.null <- glmer(G_Speech ~ 1 + (1 + M_Gesture |
  ↪ Participant), data = data_Clarify, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

G_Speech.M_Gesture <- glmer(G_Speech ~ M_Gesture + (1 + M_
  ↪ Gesture |Participant), data = data_Clarify, family =
  ↪ binomial, control = glmerControl(optimizer = "bobyqa"
  ↪ ))

anova(G_Speech.null, G_Speech.M_Gesture)
Data: data_Clarify
Models:
  G_Speech.null: G_Speech ~ 1 + (1 + M_Gesture |
    ↪ Participant)
  G_Speech.M_Gesture: G_Speech ~ M_Gesture + (1 + M_Gesture |
    ↪ Participant)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Speech.null      4 444.12 459.23 -218.06   436.12
G_Speech.M_Gesture  5 438.05 456.93 -214.02   428.05 8.0734
  ↪      1   0.004492 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

summary(G_Speech.M_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ M_Gesture + (1 + M_Gesture |
  ↪ Participant)
Data: data_Clarify
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
438.0    456.9   -214.0    428.0     318

Scaled residuals:

```

```

  Min      1Q  Median      3Q      Max
-1.1841 -1.1841  0.8445  0.8445  1.4881

Random effects:
  Groups      Name      Variance Std.Dev.  Corr
Participant (Intercept) 0.000e+00 0.000e+00
M_Gesture    5.377e-14 2.319e-07  NaN
Number of obs: 323, groups:  Participant, 10

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.3379      0.1329   2.543   0.011 *
M_Gesture   -1.1329      0.2636  -4.297 1.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
M_Gesture -0.504

##
G_Speech.null <- glmer(G_Speech ~ 1 + (1|Participant), data
  ↪ = data_Clarify, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

G_Speech.Dir_Gesture <- glmer(G_Speech ~ Dir_Gesture + (1|
  ↪ Participant), data = data_Clarify, family = binomial,
  ↪ control = glmerControl(optimizer = "bobyqa"))

anova(G_Speech.null, G_Speech.Dir_Gesture)
Data: data_Clarify
Models:
  G_Speech.null: G_Speech ~ 1 + (1 | Participant)
  G_Speech.Dir_Gesture: G_Speech ~ Dir_Gesture + (1 |
  ↪ Participant)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
G_Speech.null          2 451.46 459.02 -223.73  447.46
G_Speech.Dir_Gesture  3 447.11 458.45 -220.56  441.11
  ↪ 6.3466      1    0.01176 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                 ↪ ' ' 1

summary(G_Speech.Dir_Gesture)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: G_Speech ~ Dir_Gesture + (1 | Participant)
Data: data_Clarify

```

```

Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
447.1    458.4   -220.6    441.1     320

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.0962 -1.0962  0.9122  0.9122  1.2910

Random effects:
  Groups      Name      Variance Std.Dev.
Participant (Intercept) 0          0
Number of obs: 323, groups: Participant, 10

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.1838    0.1268   1.450  0.1471
Dir_Gesture -0.6946    0.2745  -2.531  0.0114 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr)
Dir_Gesture -0.462

#####
# Level and gesture
#####
data_levelNo21 <- subset(data_level, Level != "21")

Gesture.null <- glmer(Gesture ~ 1 + (1|Task), data = data_
  ↪ levelNo21, family = binomial, control = glmerControl(
  ↪ optimizer = "bobyqa"))

Gesture.Level <- glmer(Gesture ~ Level + (1|Task), data =
  ↪ data_levelNo21, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))

anova(Gesture.null, Gesture.Level)
Data: data_levelNo21
Models:
  Gesture.null: Gesture ~ 1 + (1 | Task)
  Gesture.Level: Gesture ~ Level + (1 | Task)
Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
Gesture.null    2 2459.1 2470.3 -1227.5    2455.1
Gesture.Level 11 2456.2 2517.8 -1217.1    2434.2 20.894
  ↪          9    0.01313 *
---

```

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1  
 ↪ ‘ ’ 1

**summary**(Gesture.Level)

Generalized linear mixed **model** fit **by** maximum likelihood (  
 ↪ Laplace Approximation) ['glmerMod']

Family: **binomial** ( logit )

Formula: Gesture ~ Level + (1 | Task)

Data: **data\_levelNo21**

Control: glmerControl(optimizer = "bobyqa")

AIC	BIC	logLik	<b>deviance</b>	<b>df.resid</b>
2456.2	2517.8	-1217.1	2434.2	1989

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-1.7424	-0.7749	-0.5050	0.9343	2.6132

Random **effects**:

Groups Name	Variance	Std.Dev.
Task (Intercept)	0.8314	0.9118

Number of obs: 2000, groups: Task, 8

Fixed **effects**:

Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.1904	0.3360	-0.567 0.57086
Level3	-0.2782	0.1212	-2.294 0.02178 *
Level5	-0.1558	0.1493	-1.044 0.29642
Level7	-0.3579	0.2015	-1.776 0.07569 .
Level9	-0.8011	0.2734	-2.931 0.00338 **
Level11	-0.8011	0.3751	-2.136 0.03268 *
Level13	0.2243	0.4460	0.503 0.61504
Level15	0.4258	0.5088	0.837 0.40274
Level17	0.3359	0.5726	0.587 0.55738
Level19	-1.1642	0.8009	-1.454 0.14603

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1  
 ↪ ‘ ’ 1

Correlation of Fixed Effects:

(Intr)	Level3	Level5	Level7	Level9	Levl11	Levl13	Levl15
↪ Levl17							
Level3	-0.163						
Level5	-0.133	0.411					
Level7	-0.092	0.309	0.263				
Level9	-0.064	0.230	0.195	0.150			
Level11	-0.050	0.170	0.149	0.113	0.086		
Level13	-0.038	0.141	0.125	0.095	0.077	0.056	
Level15	-0.034	0.124	0.109	0.083	0.066	0.049	0.045
Level17	-0.030	0.109	0.098	0.074	0.061	0.044	0.044

```

↪ 0.035
Level19 -0.021  0.078  0.071  0.054  0.044  0.032  0.032
↪ 0.026  0.026

```

```
lsmeans(Gesture.Level, pairwise~Level, adjust ="tukey")
```

```
$lsmeans
```

Level	lsmean	SE	df	asyp.LCL	asyp.UCL
1	-0.19042306	0.3359216	NA	-0.8488174	0.46797127
3	-0.46858644	0.3380768	NA	-1.1312048	0.19403188
5	-0.34626368	0.3490176	NA	-1.0303256	0.33779821
7	-0.54828440	0.3754902	NA	-1.2842317	0.18766289
9	-0.99150650	0.4193849	NA	-1.8134858	-0.16952723
11	-0.99156935	0.4907307	NA	-1.9533838	-0.02975494
13	0.03387582	0.5479351	NA	-1.0400573	1.10780893
15	0.23534066	0.6001992	NA	-0.9410282	1.41170952
17	0.14550244	0.6552763	NA	-1.1388154	1.42982032
19	-1.35464014	0.8619989	NA	-3.0441270	0.33484676
21	-13.60222416	16.2602183	NA	-45.4716665	18.26721816

Results are given **on** the logit (not the **response**) **scale**.  
Confidence level used: 0.95

```
$contrasts
```

contrast	estimate	SE	df	z.ratio	p.value
1 - 3	2.781634e-01	0.1212461	NA	2.294	0.4378
1 - 5	1.558406e-01	0.1492465	NA	1.044	0.9941
1 - 7	3.578613e-01	0.2014540	NA	1.776	0.7944
1 - 9	8.010834e-01	0.2733703	NA	2.930	0.1135
1 - 11	8.011463e-01	0.3750147	NA	2.136	0.5509
1 - 13	-2.242989e-01	0.4458684	NA	-0.503	1.0000
1 - 15	-4.257637e-01	0.5088421	NA	-0.837	0.9991
1 - 17	-3.359255e-01	0.5725781	NA	-0.587	1.0000
1 - 19	1.164217e+00	0.8008857	NA	1.454	0.9346
1 - 21	1.341180e+01	16.2563261	NA	0.825	0.9992
3 - 5	-1.223228e-01	0.1486797	NA	-0.823	0.9992
3 - 7	7.969796e-02	0.2004961	NA	0.398	1.0000
3 - 9	5.229201e-01	0.2724007	NA	1.920	0.7045
3 - 11	5.229829e-01	0.3740428	NA	1.398	0.9492
3 - 13	-5.024623e-01	0.4453130	NA	-1.128	0.9891
3 - 15	-7.039271e-01	0.5082300	NA	-1.385	0.9523
3 - 17	-6.140889e-01	0.5721938	NA	-1.073	0.9926
3 - 19	8.860537e-01	0.8006092	NA	1.107	0.9906
3 - 21	1.313364e+01	16.2568771	NA	0.808	0.9993
5 - 7	2.020207e-01	0.2169169	NA	0.931	0.9977
5 - 9	6.452428e-01	0.2847284	NA	2.266	0.4574
5 - 11	6.453057e-01	0.3823552	NA	1.688	0.8423
5 - 13	-3.801395e-01	0.4521237	NA	-0.841	0.9990
5 - 15	-5.816043e-01	0.5144160	NA	-1.131	0.9889

5 - 17	-4.917661e-01	0.5774036	NA	-0.852	0.9989
5 - 19	1.008376e+00	0.8041145	NA	1.254	0.9761
5 - 21	1.325596e+01	16.2571247	NA	0.815	0.9993
7 - 9	4.432221e-01	0.3143411	NA	1.410	0.9463
7 - 11	4.432850e-01	0.4052135	NA	1.094	0.9914
7 - 13	-5.821602e-01	0.4715033	NA	-1.235	0.9786
7 - 15	-7.836251e-01	0.5314824	NA	-1.474	0.9284
7 - 17	-6.937868e-01	0.5927061	NA	-1.171	0.9856
7 - 19	8.063557e-01	0.8151854	NA	0.989	0.9962
7 - 21	1.305394e+01	16.2574037	NA	0.803	0.9994
9 - 11	6.284997e-05	0.4447109	NA	0.000	1.0000
9 - 13	-1.025382e+00	0.5047464	NA	-2.031	0.6267
9 - 15	-1.226847e+00	0.5615210	NA	-2.185	0.5157
9 - 17	-1.137009e+00	0.6193508	NA	-1.836	0.7588
9 - 19	3.631336e-01	0.8347701	NA	0.435	1.0000
9 - 21	1.261072e+01	16.2583338	NA	0.776	0.9995
11 - 13	-1.025445e+00	0.5663459	NA	-1.811	0.7742
11 - 15	-1.226910e+00	0.6171709	NA	-1.988	0.6575
11 - 17	-1.137072e+00	0.6706427	NA	-1.695	0.8383
11 - 19	3.630708e-01	0.8733596	NA	0.416	1.0000
11 - 21	1.261065e+01	16.2608901	NA	0.776	0.9995
13 - 15	-2.014648e-01	0.6613342	NA	-0.305	1.0000
13 - 17	-1.116266e-01	0.7102710	NA	-0.157	1.0000
13 - 19	1.388516e+00	0.9040355	NA	1.536	0.9079
13 - 21	1.363610e+01	16.2624572	NA	0.839	0.9991
15 - 17	8.983822e-02	0.7523403	NA	0.119	1.0000
15 - 19	1.589981e+00	0.9374742	NA	1.696	0.8381
15 - 21	1.383756e+01	16.2665267	NA	0.851	0.9989
17 - 19	1.500143e+00	0.9722964	NA	1.543	0.9054
17 - 21	1.374773e+01	16.2671564	NA	0.845	0.9990
19 - 21	1.224758e+01	16.2783689	NA	0.752	0.9996

Results are given **on the log (not the response) scale.**

P value adjustment: tukey method **for** comparing a **family** of

↪ 11 estimates

Tests are performed **on the log scale**

```
#####
# Level and M_Gesture
#####
```

```
M_Gesture.null <- glmer(M_Gesture ~ 1 + (1|Task), data =
  ↪ data_levelNo21, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))
```

```
M_Gesture.Level <- glmer(M_Gesture ~ Level + (1|Task), data
  ↪ = data_levelNo21, family = binomial, control =
  ↪ glmerControl(optimizer = "bobyqa"))
```

```
anova(M_Gesture.null, M_Gesture.Level)
```



```

Data: data_levelNo21
Models:
  M_Gesture.null: M_Gesture ~ 1 + (1 | Task)
M_Gesture.Level: M_Gesture ~ Level + (1 | Task)
Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
M_Gesture.null    2 2064.3 2075.5 -1030.1    2060.3
M_Gesture.Level 11 2065.1 2126.7 -1021.5    2043.1 17.189
  ↪      9    0.04584 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪  '.' 1

summary(M_Gesture.Level)
Generalized linear mixed model fit by maximum likelihood (
  ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: M_Gesture ~ Level + (1 | Task)
Data: data_levelNo21
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC  logLik deviance df.resid
2065.1    2126.7 -1021.5    2043.1    1989

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.3337 -0.6383 -0.3497 -0.2153  4.6438

Random effects:
  Groups Name      Variance Std.Dev.
Task (Intercept) 0.685    0.8276
Number of obs: 2000, groups: Task, 8

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.84347    0.30924  -2.728  0.00638 **
  Level3     -0.28364    0.13215  -2.146  0.03185 *
  Level5     -0.34537    0.16839  -2.051  0.04026 *
  Level7     -0.36358    0.23117  -1.573  0.11578
Level9      -0.77746    0.32338  -2.404  0.01621 *
  Level11    -0.96946    0.49371  -1.964  0.04957 *
  Level13    -0.23762    0.49853  -0.477  0.63362
Level15     -0.01705    0.55738  -0.031  0.97559
Level17     -1.81809    1.04839  -1.734  0.08289 .
Level19     -0.67724    0.80504  -0.841  0.40021
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪  '.' 1

Correlation of Fixed Effects:
  (Intr) Level3 Level5 Level7 Level9 Lev11 Lev13 Lev15

```

```

    ↪ Level17
Level3  -0.183
Level5  -0.146  0.371
Level7  -0.100  0.276  0.228
Level9  -0.067  0.200  0.167  0.129
Level11 -0.049  0.133  0.115  0.087  0.066
Level13 -0.042  0.130  0.114  0.087  0.069  0.046
Level15 -0.038  0.117  0.100  0.076  0.061  0.041  0.044
Level17 -0.019  0.062  0.054  0.041  0.033  0.022  0.025
    ↪ 0.022
Level19 -0.026  0.081  0.073  0.056  0.045  0.030  0.034
    ↪ 0.028  0.016

lsmeans(M_Gesture.Level, pairwise~Level, adjust ="tukey")
#####
# Level and Dir_Gesture
#####

Dir_Gesture.null <- glmer(Dir_Gesture ~ 1 + (1|Task), data
    ↪ = data_levelNo21, family = binomial, control =
    ↪ glmerControl(optimizer = "bobyqa"))

Dir_Gesture.Level <- glmer(Dir_Gesture ~ Level + (1|Task),
    ↪ data = data_levelNo21, family = binomial, control =
    ↪ glmerControl(optimizer = "bobyqa"))

anova(Dir_Gesture.null, Dir_Gesture.Level)
Data: data_levelNo21
Models:
  Dir_Gesture.null: Dir_Gesture ~ 1 + (1 | Task)
  Dir_Gesture.Level: Dir_Gesture ~ Level + (1 | Task)
Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
Dir_Gesture.null    2 1949.9 1961.1 -972.93   1945.9
Dir_Gesture.Level  11 1949.5 2011.1 -963.72   1927.5 18.413
    ↪      9    0.03067 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

summary(Dir_Gesture.Level)
Generalized linear mixed model fit by maximum likelihood (
    ↪ Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: Dir_Gesture ~ Level + (1 | Task)
Data: data_levelNo21
Control: glmerControl(optimizer = "bobyqa")

AIC      BIC    logLik deviance df.resid
1949.4    2011.1   -963.7   1927.4     1989

```

```

Scaled residuals:
  Min      1Q  Median      3Q      Max
-1.2934 -0.5651 -0.3592 -0.2438  6.5241

Random effects:
  Groups Name      Variance Std.Dev.
Task  (Intercept) 0.6304   0.794
Number of obs: 2000, groups: Task, 8

Fixed effects:
  Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.9477      0.2984  -3.176  0.00149 **
  Level3      -0.2183      0.1362  -1.603  0.10886
Level5       -0.2157      0.1733  -1.244  0.21340
Level7       -0.3853      0.2477  -1.555  0.11984
Level9       -0.9059      0.3714  -2.439  0.01472 *
  Level11     -1.6809      0.7328  -2.294  0.02180 *
  Level13     -0.3410      0.5701  -0.598  0.54979
Level15       0.3610      0.5553   0.650  0.51566
Level17     -1.3619      1.0494  -1.298  0.19433
Level19     -0.1740      0.8035  -0.216  0.82860
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ' ' 1
  ↪ ' ' 1

Correlation of Fixed Effects:
  (Intr) Level3 Level5 Level7 Level9 Lev11 Lev13 Lev15
  ↪ Lev17
Level3 -0.197
Level5 -0.159  0.372
Level7 -0.105  0.267  0.221
Level9 -0.065  0.180  0.149  0.111
Level11 -0.037  0.092  0.080  0.058  0.041
Level13 -0.041  0.117  0.102  0.075  0.056  0.029
Level15 -0.043  0.121  0.105  0.077  0.057  0.030  0.042
Level17 -0.022  0.063  0.056  0.041  0.030  0.015  0.023
  ↪ 0.023
Level19 -0.029  0.083  0.076  0.056  0.041  0.022  0.033
  ↪ 0.032  0.018

#####
# G_Speech and Level
#####

G_Speech.null <- glmer(G_Speech ~ 1 + (1|Task), data = data
  ↪ _levelNo21, family = binomial, control = glmerControl
  ↪ (optimizer = "bobyqa"))

G_Speech.Level <- glmer(G_Speech ~ Level + (1|Task), data =

```

```
↪ data_levelNo21, family = binomial, control =
↪ glmerControl(optimizer = "bobyqa"))
```

```
anova(G_Speech.null, G_Speech.Level)
```

```
Data: data_levelNo21
```

```
Models:
```

```
G_Speech.null: G_Speech ~ 1 + (1 | Task)
```

```
G_Speech.Level: G_Speech ~ Level + (1 | Task)
```

```
Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
```

```
G_Speech.null      2 2735.6 2746.8 -1365.8 2731.6
```

```
G_Speech.Level 11 2735.0 2796.6 -1356.5 2713.0 18.626
```

```
↪      9      0.02857 *
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
↪  ' ' 1
```

```
summary(G_Speech.Level)
```

```
Generalized linear mixed model fit by maximum likelihood (
```

```
↪ Laplace Approximation) ['glmerMod']
```

```
Family: binomial (logit)
```

```
Formula: G_Speech ~ Level + (1 | Task)
```

```
Data: data_levelNo21
```

```
Control: glmerControl(optimizer = "bobyqa")
```

```
AIC      BIC  logLik deviance df.resid
2735.0    2796.6 -1356.5 2713.0    1989
```

```
Scaled residuals:
```

```
  Min      1Q  Median      3Q      Max
-1.5026 -1.0289  0.7408  0.9113  1.5298
```

```
Random effects:
```

```
Groups Name          Variance Std.Dev.
```

```
Task (Intercept) 0.1998  0.447
```

```
Number of obs: 2000, groups: Task, 8
```

```
Fixed effects:
```

```
Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.1534      0.1804  0.850  0.3953
Level3      0.1380      0.1139  1.212  0.2255
Level5     -0.1074      0.1396 -0.769  0.4420
Level7      0.2524      0.1868  1.351  0.1766
Level9     -0.2574      0.2335 -1.103  0.2703
Level11    -0.6629      0.3077 -2.155  0.0312 *
Level13    -0.4580      0.4459 -1.027  0.3043
Level15    -0.7438      0.5181 -1.436  0.1511
Level17    -0.8764      0.6103 -1.436  0.1510
Level19    -0.4771      0.6553 -0.728  0.4666
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

↪ ' ' 1

Correlation of Fixed Effects:

```
(Intr) Level3 Level5 Level7 Level9 Lev111 Lev113 Lev115
↪ Level17
Level3 -0.290
Level5 -0.237 0.420
Level7 -0.169 0.321 0.273
Level9 -0.129 0.260 0.222 0.173
Level111 -0.105 0.202 0.176 0.136 0.112
Level113 -0.066 0.134 0.118 0.091 0.079 0.058
Level115 -0.057 0.117 0.101 0.078 0.067 0.051 0.039
Level117 -0.048 0.098 0.087 0.067 0.058 0.042 0.036
↪ 0.029
Level119 -0.044 0.092 0.083 0.064 0.056 0.041 0.035
↪ 0.028 0.026
```

ScatterPlots

```
par(mfrow=c(2,2))
scatter2 <- ggplot(data_levelNo21, aes(Level, Gesture))

scatter2 + stat_summary(fun.y = mean, geom = "point") +
↪ labs(x = "Level", y = "Gesture") + stat_summary(fun.
↪ data = mean_cl_boot, geom = "errorbar", width = 0.2)

scatterMGest <- ggplot(data_levelNo21, aes(Level, M_Gesture
↪ ))

scatterMGest + stat_summary(fun.y = mean, geom = "point") +
↪ labs(x = "Level", y = "Manner_in_Gesture") + stat_
↪ summary(fun.data = mean_cl_boot, geom = "errorbar",
↪ width = 0.2)

scatterDirGest <- ggplot(data_levelNo21, aes(Level, Dir_
↪ Gesture))

scatterDirGest + stat_summary(fun.y = mean, geom = "point") +
↪ + labs(x = "Level", y = "Direction_in_Gesture") +
↪ stat_summary(fun.data = mean_cl_boot, geom = "
↪ errorbar", width = 0.2)

scatterGSpeech <- ggplot(data_levelNo21, aes(Level, G_
↪ Speech))

scatterGSpeech + stat_summary(fun.y = mean, geom = "point") +
↪ + labs(x = "Level", y = "Ground_in_Speech") + stat_
↪ summary(fun.data = mean_cl_boot, geom = "errorbar",
↪ width = 0.2)
```

```
library(ggplot2)
library(grid)
library(gridExtra)

grid.arrange(scatter2, scatterGSpeech, scatterMGest,
  ↪ scatterDirGest, ncol = 2, main = "Main_title")

multiplot(scatter2, scatterGSpeech, scatterMGest,
  ↪ scatterDirGest, cols=2)
attach(mtcars)
```

## Eye tracking R script

```

#Calculate proportions

e<-ddply(d,.(Participant,MediaName, condition, Timebin),
  ↪ transform,prop=FixationCount/sum(FixationCount))

write.csv(e, "etPropFinal.csv")

e <- read.csv("etPropFinal.csv", header = T)

e$condition <- as.factor(e$condition)

#Target advantage

#Target advantage is the proprtion of looks to target minus
  ↪ the proportion of looks to every other item in the
  ↪ display. Character was excluded from this analysis

#Remove scores for hits to target and hits to character

f <- e[e$AOI != "AOI.Target.Hit",]

f <- f[f$AOI != "AOI.Character.Hit",]

#2. create a column containing the average proportion of
  ↪ looks to distractors per item per participant

f<-ddply(f,.(Participant, MediaName, condition, Timebin),
  ↪ transform,avDisProp=mean(prop))

#3 create a df only including target prop

g <- e[e$AOI == "AOI.Target.Hit",]

#4. create a column containing the average proportion of
  ↪ looks to targets per item per participant

h <- ddply(g,.(MediaName, Participant, Timebin),transform,
  ↪ TargetProp=mean(prop))

#5.create a new df containing only those columns of
  ↪ interest including the new avDisProp column

i <- f[, c("Participant", "MediaName", "condition", "
  ↪ Timebin", "gesture", "speech", "avDisProp")]

i$avDisProp <- ifelse(i$avDisProp >0, 1, 0)

i <- unique(i[c("Participant", "MediaName", "condition", "

```

```

  ↪ Timebin", "gesture", "speech", "avDisProp"))])
j <- h[, c("Participant", "MediaName", "condition", "
  ↪ Timebin", "gesture", "speech", "TargetProp")]
df1 <- merge(i, j, all=T)
df1$targetAdvantage <- df1$TargetProp - df1$avDisProp
write.csv(df1, "targetAdvantage.csv")
df1 <- read.csv("targetAdvantage.csv")

#plotting proportions

library(ggplot2)
e_con1 <- subset(e, condition == "condition1")
scatter1 <- ggplot(e_con1, aes(Timebin, prop, colour = AOI)
  ↪ )

e_con2 <- subset(e, condition == "condition2")
scatter2 <- ggplot(e_con2, aes(Timebin, prop, colour = AOI)
  ↪ )

e_con3 <- subset(e, condition == "condition3")
scatter3 <- ggplot(e_con3, aes(Timebin, prop, colour = AOI)
  ↪ )

e_con4 <- subset(e, condition == "condition4")
scatter4 <- ggplot(e_con4, aes(Timebin, prop, colour = AOI)
  ↪ )

scatter1 + stat_summary(fun.y = mean, geom = "point") +
  ↪ stat_summary(fun.y = mean, geom = "line", aes (group
  ↪ = AOI)) + labs(colour = "AOI") + labs(x = "Time(ms)",
  ↪ y = "Proportion", colour = "Area_of_Interest") +
  ↪ ggtitle("+_Speech;_+_Gesture") + annotate("rect",
  ↪ xmin = 5270, xmax = 5540, ymin = -0.0, ymax = 0.99,
  ↪ alpha = .2) + annotate("rect", xmin = 6950, xmax =
  ↪ 7500, ymin = -0.0, ymax = 0.99, alpha = .2) + scale_
  ↪ color_manual("Area_of_Interest", breaks=c("AOI.
  ↪ Character.Hit", "AOI.D1.Hit", "AOI.D2.Hit", "AOI.D3.
  ↪ Hit", "AOI.Target.Hit"), labels=c("Character", "D|O_
  ↪ Distractor", "D|O_Distractor", "Distractor", "Target"
  ↪ ), values = c("red", "blue", "green", "pink", "
  ↪ darkgreen")) + theme(legend.text=element_text(size
  ↪ =12), axis.text=element_text(size=12), axis.title=
  ↪ element_text(size=14,face="bold"), legend.title=
  ↪ element_text(size=14), plot.title=element_text(size
  ↪ =16)) + annotate("text", x = 5000, y = 0.6, label = "
  ↪ mannerDOP") + annotate("text", x = 8000, y = 0.85,

```



```
→ label = "groundDOP")
```

```
scatter2 + stat_summary(fun.y = mean, geom = "point") +
→ stat_summary(fun.y = mean, geom = "line", aes (group
→ = AOI)) + labs(colour = "AOI") + labs(x = "Time(ms)",
→ y = "Proportion", colour = "condition2")+ ggtitle("-<
→ Speech;-<
→ Gesture")+ annotate("rect", xmin = 5270,
→ xmax = 5540, ymin = 0, ymax = 0.99, alpha = .2)+
→ annotate("rect", xmin = 6950, xmax = 7500, ymin =
→ -0.0, ymax = 0.99, alpha = .2) + scale_color_manual("
→ Area<
→ of<
→ Interest", breaks=c("AOI.Character.Hit", "AOI.
→ D1.Hit", "AOI.D2.Hit", "AOI.D3.Hit", "AOI.Target.Hit"
→ ), labels=c("Character", "D|O<
→ Distractor", "D|O<
→ Distractor", "Distractor", "Target"), values = c("red
→ ", "blue", "green", "pink", "darkgreen")) + theme(
→ legend.text=element_text(size=12), axis.text=element_
→ text(size=12), axis.title=element_text(size=14,face="
→ bold"), legend.title=element_text(size=14), plot.
→ title=element_text(size=16)) + annotate("text", x =
→ 5000, y = 0.6, label = "mannerDOP") + annotate("text"
→ , x = 8000, y = 0.85, label = "groundDOP")
```

```
scatter3 + stat_summary(fun.y = mean, geom = "point") +
→ stat_summary(fun.y = mean, geom = "line", aes (group
→ = AOI)) + labs(colour = "AOI") + labs(x = "Time(ms)",
→ y = "Proportion", colour = "group")+ ggtitle("-<
→ Speech;<
→ +<
→ Gesture")+ annotate("rect", xmin = 5270,
→ xmax = 5540, ymin = 0, ymax = 0.99, alpha = .2)+
→ annotate("rect", xmin = 6950, xmax = 7500, ymin =
→ -0.0, ymax = 0.99, alpha = .2) + scale_color_manual("
→ Area<
→ of<
→ Interest", breaks=c("AOI.Character.Hit", "AOI.
→ D1.Hit", "AOI.D2.Hit", "AOI.D3.Hit", "AOI.Target.Hit"
→ ), labels=c("Character", "D|O<
→ Distractor", "D|O<
→ Distractor", "Distractor", "Target"), values = c("red
→ ", "blue", "green", "pink", "darkgreen")) + theme(
→ legend.text=element_text(size=12), axis.text=element_
→ text(size=12), axis.title=element_text(size=14,face="
→ bold"), legend.title=element_text(size=14), plot.
→ title=element_text(size=16)) + annotate("text", x =
→ 5000, y = 0.6, label = "mannerDOP") + annotate("text"
→ , x = 8000, y = 0.85, label = "groundDOP")
```

```
scatter4 + stat_summary(fun.y = mean, geom = "point") +
→ stat_summary(fun.y = mean, geom = "line", aes (group
→ = AOI)) + labs(colour = "AOI") + labs(x = "Time(ms)",
→ y = "Proportion", colour = "group") + ggtitle("-<
```

```

↪ Speech;_Gesture")+ annotate("rect", xmin = 5270,
↪ xmax = 5540, ymin = 0, ymax = 0.99, alpha = .2)+
↪ annotate("rect", xmin = 6950, xmax = 7500, ymin =
↪ -0.0, ymax = 0.99, alpha = .2) + scale_color_manual("
↪ Area_of_Interest", breaks=c("AOI.Character.Hit", "AOI.
↪ D1.Hit", "AOI.D2.Hit", "AOI.D3.Hit", "AOI.Target.Hit"
↪ ), labels=c("Character", "D|O_Distractor", "D|O_
↪ Distractor", "Distractor", "Target"), values = c("red
↪ ", "blue", "green", "pink", "darkgreen")) + theme(
↪ legend.text=element_text(size=12), axis.text=element_
↪ text(size=12), axis.title=element_text(size=14,face="
↪ bold"), legend.title=element_text(size=14), plot.
↪ title=element_text(size=16)) + annotate("text", x =
↪ 5000, y = 0.6, label = "mannerDOP") + annotate("text"
↪ , x = 8000, y = 0.85, label = "groundDOP")

```

*#Plotting Target advantage*

```

scatterTa <- ggplot(df1, aes(Timebin, targetAdvantage,
↪ colour = condition))

```

```

scatterTa + stat_summary(fun.y = mean, geom = "point") +
↪ stat_summary(fun.y = mean, geom = "line", aes (group
↪ = condition)) + labs(x = "Time(ms)", y = "Targert_
↪ Advantage", colour = "condition") + ggtitle("Target_
↪ Advantage") + annotate("rect", xmin = 5270, xmax =
↪ 5540, ymin = -0.50, ymax = 0.75, alpha = .2)+
↪ annotate("rect", xmin = 6950, xmax = 7500, ymin =
↪ -0.5, ymax = 0.75, alpha = .2) + scale_color_manual("
↪ Condition", breaks=c("condition1", "condition2", "
↪ condition3", "condition4"), labels=c("+Speech;_+
↪ Gesture", "+Speech;_Gesture", "-Speech;_+Gesture", "
↪ -Speech;_Gesture"), values = c("red", "blue", "green
↪ ", "darkgreen")) + theme(legend.text=element_text(
↪ size=12), axis.text=element_text(size=12), axis.title
↪ =element_text(size=14,face="bold"), legend.title=
↪ element_text(size=14), plot.title=element_text(size
↪ =16)) + annotate("text", x = 6100, y = 0.6, label = "
↪ mannerDOP") + annotate("text", x = 8000, y = 0.00,
↪ label = "groundDOP")

```

*#5250-5550*

```

r<- subset(df1, Timebin >= 5250 & Timebin <= 5550)

```

```

r.null <- lmer(targetAdvantage ~ 1 + (1+speech+gesture|
↪ Participant) + (1|MediaName), data = r, REML=FALSE)

```

```

r.speech <- lmer(targetAdvantage ~ speech + (1+speech+

```

```

  ↪ gesture|Participant) + (1|MediaName), data = r, REML=
  ↪ FALSE)

r.gesture <- lmer(targetAdvantage ~ gesture + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = r, REML=
  ↪ FALSE)

r.mannerFull <- lmer(targetAdvantage ~ speech*gesture + (1+
  ↪ speech+gesture|Participant) + (1|MediaName), data = r
  ↪ , REML=FALSE)

r.mannerSub <- lmer(targetAdvantage ~ speech + gesture +
  ↪ (1+speech+gesture|Participant) + (1|MediaName), data
  ↪ = r, REML=FALSE)

anova(r.null, r.gesture)
Data: r
Models:
  r.null: targetAdvantage ~ 1 + (1 + speech + gesture |
  ↪ Participant) +
  r.null:      (1 | MediaName)
r.gesture: targetAdvantage ~ gesture + (1 + speech +
  ↪ gesture | Participant) +
  r.gesture:      (1 | MediaName)
Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
r.null      9 13531 13606 -6756.6    13513
r.gesture  10 13533 13617 -6756.6    13513 0.0585      1
  ↪      0.809

anova(r.null, r.speech)
Data: r
Models:
  r.null: targetAdvantage ~ 1 + (1 + speech + gesture |
  ↪ Participant) +
  r.null:      (1 | MediaName)
r.speech: targetAdvantage ~ speech + (1 + speech + gesture
  ↪ | Participant) +
  r.speech:      (1 | MediaName)
Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
r.null      9 13531 13606 -6756.6    13513
r.speech  10 13533 13616 -6756.4    13513 0.5115      1
  ↪      0.4745

anova(r.mannerSub, r.mannerFull)
Data: r
Models:
  r.mannerSub: targetAdvantage ~ speech + gesture + (1 +
  ↪ speech + gesture | r.mannerSub:      Participant) +
  ↪ (1 | MediaName)
r.mannerFull: targetAdvantage ~ speech * gesture + (1 +

```

```

    ↪ speech + gesture | r.mannerFull: Participant) + (1 |
    ↪ MediaName)
Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
r.mannerSub  11 13535 13626 -6756.3    13513
r.mannerFull 12 13537 13637 -6756.3    13513 0.0103      1
    ↪      0.9192

summary(r.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: targetAdvantage ~ speech * gesture + (1 + speech +
    ↪ gesture | Participant) + (1 | MediaName)
Data: r

AIC      BIC    logLik deviance df.resid
13536.7  13636.8 -6756.3  13512.7    31056

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.7338 -0.0408  0.1127  0.2704  4.5458

Random effects:
  Groups      Name          Variance Std.Dev. Corr
MediaName (Intercept)  0.003153 0.05615
Participant (Intercept) 0.003129 0.05594
speechwiggly 0.002704 0.05200 -0.03
gesturepresent 0.002108 0.04592 -0.28 0.05
Residual              0.088806 0.29800
Number of obs: 31068, groups: MediaName, 158; Participant,
    ↪ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)          -0.055936  0.013572  -4.122
speechwiggly         -0.008459  0.016273  -0.520
gesturepresent       -0.002234  0.015669  -0.143
speechwiggly:gesturepresent -0.001941  0.019130  -0.101

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.433
gesturprsnt -0.536  0.376
spchwggly:g  0.356 -0.594 -0.610

lsmeans(r.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL
    ↪ asymp.UCL
dotted absent -0.05593626 0.01357164 NA -0.08253618
    ↪ -0.02933635

```

```
wiggly absent -0.06439511 0.01604844 NA -0.09584947
  ↪ -0.03294076
dotted present -0.05816978 0.01420026 NA -0.08600177
  ↪ -0.03033779
wiggly present -0.06856943 0.01680180 NA -0.10150034
  ↪ -0.03563852
```

Confidence level used: 0.95

### \$contrasts

contrast	estimate	SE	df
↪ z.ratio p.value			
dotted,absent - wiggly,absent	0.008458849	0.01627268	NA
↪ 0.520 0.9544			
dotted,absent - dotted,present	0.002233517	0.01566882	NA
↪ 0.143 0.9990			
dotted,absent - wiggly,present	0.012633167	0.01824056	NA
↪ 0.693 0.9000			
wiggly,absent - dotted,present	-0.006225331	0.01784915	NA
↪ -0.349 0.9854			
wiggly,absent - wiggly,present	0.004174318	0.01568477	NA
↪ 0.266 0.9934			
dotted,present - wiggly,present	0.010399650	0.01615372	NA
↪ 0.644 0.9177			

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

```
#5550 - 5850
```

```
s<- subset(df1, Timebin >= 5550 & Timebin <= 5850)
```

```
s.null <- lmer(targetAdvantage ~ 1 + (1+gesture+speech|
  ↪ Participant) + (1|MediaName), data = s, REML=FALSE)
```

```
s.speech <- lmer(targetAdvantage ~ speech + (1+gesture+
  ↪ speech|Participant) + (1|MediaName), data = s, REML=
  ↪ FALSE)
```

```
s.gesture <- lmer(targetAdvantage ~ gesture + (1+gesture+
  ↪ speech|Participant) + (1|MediaName), data = s, REML=
  ↪ FALSE)
```

```
s.mannerFull <- lmer(targetAdvantage ~ speech*gesture + (1+
  ↪ gesture+speech|Participant) + (1|MediaName), data = s
  ↪ , REML=FALSE)
```

```
s.mannerSub <- lmer(targetAdvantage ~ speech + gesture +
  ↪ (1+gesture+speech|Participant) + (1|MediaName), data
  ↪ = s, REML=FALSE)
```

```

anova(s.null, s.gesture)
Data: s
Models:
  s.null: targetAdvantage ~ 1 + (1 + gesture + speech |
    ↪ Participant) +
  s.null:      (1 | MediaName)
s.gesture: targetAdvantage ~ gesture + (1 + gesture +
    ↪ speech | Participant) +
  s.gesture:      (1 | MediaName)
Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
s.null      9 39850 39925 -19916      39832
s.gesture  10 39851 39935 -19916      39831 0.5642      1
    ↪ 0.4526

anova(s.null, s.speech)
Data: s
Models:
  s.null: targetAdvantage ~ 1 + (1 + gesture + speech |
    ↪ Participant) +
  s.null:      (1 | MediaName)
s.speech: targetAdvantage ~ speech + (1 + gesture + speech
    ↪ | Participant) +
  s.speech:      (1 | MediaName)
      Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>
    ↪ Chisq)
s.null      9 39850 39925 -19916      39832
s.speech  10 39850 39934 -19915      39830 1.8052      1
    ↪ 0.1791

anova(s.mannerSub, s.mannerFull)
Data: s
Models:
  s.mannerSub: targetAdvantage ~ speech + gesture + (1 +
    ↪ gesture + speech | s.mannerSub: Participant) + (1 |
    ↪ MediaName)
s.mannerFull: targetAdvantage ~ speech * gesture + (1 +
    ↪ gesture + speech | s.mannerFull: Participant) + (1 |
    ↪ MediaName)
Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
s.mannerSub  11 39852 39944 -19915      39830
s.mannerFull 12 39854 39954 -19915      39830      0      1
    ↪      0.9955

lsmeans(s.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL
    ↪ asymp.UCL
dotted absent -0.10005079 0.02024021 NA -0.1397209
    ↪ -0.060380705

```

```
wiggly absent -0.06811331 0.02743010 NA -0.1218753
  ↪ -0.014351309
dotted present -0.08776619 0.02093122 NA -0.1287906
  ↪ -0.046741751
wiggly present -0.05599765 0.02692647 NA -0.1087726
  ↪ -0.003222742
```

Confidence level used: 0.95

### \$contrasts

contrast	estimate	SE	df	z
↪ .ratio				p.value
dotted,absent - wiggly,absent	-0.03193748	0.02839562	NA	
↪	-1.125	0.6742		
dotted,absent - dotted,present	-0.01228460	0.02354223	NA	
↪	-0.522	0.9539		
dotted,absent - wiggly,present	-0.04405314	0.02911315	NA	
↪	-1.513	0.4295		
wiggly,absent - dotted,present	0.01965288	0.03098084	NA	
↪	0.634	0.9210		
wiggly,absent - wiggly,present	-0.01211566	0.02355554	NA	
↪	-0.514	0.9557		
dotted,present - wiggly,present	-0.03176854	0.02820404	NA	
↪	-1.126	0.6732		

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

### summary(s.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: targetAdvantage ~ speech \* gesture + (1 + gesture  
 ↪ + speech | Participant) + (1 | MediaName)  
 Data: s

AIC	BIC	logLik	deviance	df.resid
39853.8	39954.1	-19914.9	39829.8	31445

### Scaled residuals:

Min	1Q	Median	3Q	Max
-2.7305	-0.1419	0.0839	0.2849	3.2126

### Random effects:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.007884	0.08879	
Participant	(Intercept)	0.006162	0.07850	
gesturepresent		0.003424	0.05852	-0.24
speechwiggly		0.011954	0.10934	0.00
Residual		0.203823	0.45147	

Number of obs: 31457, groups: MediaName, 158; Participant,  
 ↪ 35



Fixed **effects**:

	Estimate	Std. Error	<b>t</b> value
(Intercept)	-0.1000508	0.0202402	-4.943
speechwiggly	0.0319375	0.0283956	1.125
gesturepresent	0.0122846	0.0235422	0.522
speechwiggly:gesturepresent	-0.0001689	0.0300581	-0.006

## Correlation of Fixed Effects:

	(Intr)	spchwg	gstrpr
speechwggly	-0.403		
gesturprsnt	-0.552	0.300	
spchwggly:g	0.376	-0.536	-0.638

#5850-6150

**t** <- subset(df1, Timebin >= 5850 & Timebin <= 6150)**t.null** <- lmer(targetAdvantage ~ 1 + (1+gesture+speech|  
↪ Participant) + (1|MediaName), **data** = **t**, REML=FALSE)**t.speech** <- lmer(targetAdvantage ~ speech + (1+gesture+  
↪ speech|Participant) + (1|MediaName), **data** = **t**, REML=  
↪ FALSE)**t.gesture** <- lmer(targetAdvantage ~ gesture + (1+gesture+  
↪ speech|Participant) + (1|MediaName), **data** = **t**, REML=  
↪ FALSE)**t.mannerFull** <- lmer(targetAdvantage ~ speech\*gesture + (1+  
↪ gesture+speech|Participant) + (1|MediaName), **data** = **t**  
↪ , REML=FALSE)**t.mannerSub** <- lmer(targetAdvantage ~ speech + gesture +  
↪ (1+gesture+speech|Participant) + (1|MediaName), **data**  
↪ = **t**, REML=FALSE)anova(**t.null**, **t.gesture**)Data: **t**

Models:

**t.null**: targetAdvantage ~ 1 + (1 + gesture + speech |  
↪ Participant) +**t.null**: (1 | MediaName)**t.gesture**: targetAdvantage ~ gesture + (1 + gesture +  
↪ speech | Participant) + **t.gesture**: (1 | MediaName  
↪ )Df AIC BIC logLik **deviance** Chisq Chi Df Pr(>Chisq)**t.null** 9 59472 59547 -29727 59454**t.gesture** 10 59470 59553 -29725 59450 4.1528 1

↪ 0.04157 \*

---



```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                ↪ ' ' 1

anova(t.null, t.speech)
Data: t
Models:
  t.null: targetAdvantage ~ 1 + (1 + gesture + speech |
    ↪ Participant) + t.null: (1 | MediaName) t.speech:
    ↪ targetAdvantage ~ speech + (1 + gesture + speech |
    ↪ Participant) +
  t.speech:      (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
t.null     9 59472 59547 -29727   59454
t.speech  10 59457 59541 -29719   59437 16.478      1
    ↪ 4.921e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                ↪ ' ' 1

anova(t.mannerSub, t.mannerFull)
Data: t
Models:
  t.mannerSub: targetAdvantage ~ speech + gesture + (1 +
    ↪ gesture + speech | t.mannerSub: Participant) + (1 |
    ↪ MediaName)
  t.mannerFull: targetAdvantage ~ speech * gesture + (1 +
    ↪ gesture + speech | t.mannerFull: Participant) + (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
t.mannerSub  11 59447 59538 -29712   59425
t.mannerFull 12 59446 59546 -29711   59422 3.0229      1
    ↪ 0.0821 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
                ↪ ' ' 1

lsmeans(t.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean          SE df  asymp.LCL
    ↪ asymp.UCL
dotted absent -0.19592394 0.03214465 NA -0.258926297
    ↪ -0.13292157
wiggly absent  0.08021725 0.03928526 NA  0.003219545
    ↪ 0.15721495
dotted present -0.02227260 0.03159022 NA -0.084188282
    ↪ 0.03964309
wiggly present  0.16354494 0.04559412 NA  0.074182103
    ↪ 0.25290778

```

Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z
↪ .ratio p.value				
dotted,absent - wiggly,absent	-0.27614118	0.04857856	NA	
↪ -5.684	<.0001			
dotted,absent - dotted,present	-0.17365134	0.04301702	NA	
↪ -4.037	0.0003			
dotted,absent - wiggly,present	-0.35946888	0.05852432	NA	
↪ -6.142	<.0001			
wiggly,absent - dotted,present	0.10248984	0.04779385	NA	
↪ 2.144	0.1392			
wiggly,absent - wiggly,present	-0.08332769	0.04299699	NA	
↪ -1.938	0.2121			
dotted,present - wiggly,present	-0.18581754	0.04820921	NA	
↪ -3.854	0.0007			

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary(t.mannerFull)**

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: targetAdvantage ~ speech \* gesture + (1 + gesture  
 ↪ + speech | Participant) + (1 | MediaName)  
 Data: **t**

AIC	BIC	logLik	deviance	df.resid
59445.6	59545.8	-29710.8	59421.6	31221

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-2.9227	-0.5300	0.0382	0.6023	2.7522

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.02445	0.1564	
Participant	(Intercept)	0.01205	0.1098	
gesturepresent		0.01747	0.1322	-0.62
speechwiggly		0.03465	0.1861	-0.41 0.40
Residual		0.38395	0.6196	

Number of obs: 31233, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t value
(Intercept)	-0.19592	0.03214 -6.095
speechwiggly	0.27614	0.04858 5.684
gesturepresent	0.17365	0.04302 4.037

```

speechwggly:gesturepresent -0.09032    0.05171  -1.747

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.592
gesturprsnt -0.682  0.461
spchwggly:g  0.408 -0.539 -0.601

#6150-6450
u <- subset(df1, Timebin >= 6150 & Timebin <= 6450)

u.null <- lmer(targetAdvantage ~ 1 + (1+speech+gesture|
  ↪ Participant) + (1+speech+gesture|MediaName), data = u
  ↪ , REML=FALSE)

u.speech <- lmer(targetAdvantage ~ speech + (1+speech+
  ↪ gesture|Participant) + (1+speech+gesture|MediaName),
  ↪ data = u, REML=FALSE)

u.gesture <- lmer(targetAdvantage ~ gesture + (1+speech+
  ↪ gesture|Participant) + (1+speech+gesture|MediaName),
  ↪ data = u, REML=FALSE)

u.mannerFull <- lmer(targetAdvantage ~ speech*gesture + (1+
  ↪ speech+gesture|Participant) + (1+speech+gesture|
  ↪ MediaName), data = u, REML=FALSE)

u.mannerSub <- lmer(targetAdvantage ~ speech + gesture +
  ↪ (1+speech+gesture|Participant) + (1+speech+gesture|
  ↪ MediaName), data = u, REML=FALSE)

anova(u.null, u.speech)
Data: u
Models:
  u.null: targetAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) + u.null: (1 + speech + gesture |
    ↪ MediaName)
  u.speech: targetAdvantage ~ speech + (1 + speech + gesture
    ↪ | Participant) + u.speech: (1 + speech + gesture |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
u.null   14 67653 67770 -33812   67625
u.speech 15 67632 67757 -33801   67602 23.184     1
  ↪ 1.472e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

anova(u.null, u.gesture)
Data: u

```

```

Models:
  u.null: targetAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) + u.null:      (1 + speech + gesture |
    ↪ MediaName)
u.gesture: targetAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + u.gesture:      (1 + speech +
    ↪ gesture | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
u.null      14 67653 67770 -33812      67625
u.gesture  15 67637 67762 -33803      67607 17.993      1
    ↪ 2.217e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

anova(u.mannerSub, u.mannerFull)
Data: u
Models:
  u.mannerSub: targetAdvantage ~ speech + gesture + (1 +
    ↪ speech + gesture | u.mannerSub: Participant) + (1 +
    ↪ speech + gesture | MediaName)
  u.mannerFull: targetAdvantage ~ speech * gesture + (1 +
    ↪ speech + gesture | u.mannerFull: Participant) + (1
    ↪ + speech + gesture | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
u.mannerSub  16 67598 67732 -33783      67566
u.mannerFull 17 67589 67731 -33778      67555 11.175      1
    ↪ 0.000829 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

lsmeans(u.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL  asymp.
    ↪ UCL
dotted absent -0.2819291 0.04811444 NA -0.37623168
    ↪ -0.1876266
wiggly absent  0.2372860 0.05225644 NA  0.13486526
    ↪ 0.3397067
dotted present 0.1347038 0.03723797 NA  0.06171869
    ↪ 0.2076888
wiggly present 0.4104119 0.04721362 NA  0.31787492
    ↪ 0.5029489

Confidence level used: 0.95

$contrasts
contrast                                     estimate      SE df z.

```

```

  ↪ ratio p.value
dotted,absent - wiggly,absent   -0.5192151 0.06756722 NA
  ↪ -7.684 <.0001
dotted,absent - dotted,present -0.4166329 0.05661797 NA
  ↪ -7.359 <.0001
dotted,absent - wiggly,present -0.6923410 0.06672331 NA
  ↪ -10.376 <.0001
wiggly,absent - dotted,present   0.1025822 0.06132201 NA
  ↪ 1.673 0.3381
wiggly,absent - wiggly,present  -0.1731259 0.05295233 NA
  ↪ -3.269 0.0059
dotted,present - wiggly,present -0.2757082 0.05224644 NA
  ↪ -5.277 <.0001

```

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(u.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: targetAdvantage ~ speech \* gesture + (1 + speech +  
 ↪ gesture | Participant) + (1 + speech + gesture |  
 ↪ MediaName)

Data: u

AIC	BIC	logLik	deviance	df.resid
67589.2	67731.0	-33777.6	67555.2	30998

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-2.7331	-0.7076	0.1104	0.7959	2.5431

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.0704515	0.26543	
speechwiggly	0.0004681	0.02164	-1.00	
gesturepresent	0.0383625	0.19586	-0.73	0.73
Participant	(Intercept)	0.0150375	0.12263	
speechwiggly	0.0378827	0.19463	-0.28	
gesturepresent	0.0155201	0.12458	-0.43	0.19
Residual		0.5050907	0.71070	

Number of obs: 31015, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t value
(Intercept)	-0.28193	0.04811 -5.860
speechwiggly	0.51922	0.06757 7.684
gesturepresent	0.41663	0.05662 7.359
speechwiggly:gesturepresent	-0.24351	0.07136 -3.412

```

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwgly -0.638
gesturprsnt -0.759  0.524
spchwggly:g  0.545 -0.718 -0.680

#6500-6750
v <- subset(df1, Timebin >= 6450 & Timebin <= 6750)

v.null <- lmer(targetAdvantage ~ 1 + (1+speech+gesture|
  ↪ Participant) + (1|MediaName), data = v, REML=FALSE)

v.speech <- lmer(targetAdvantage ~ speech + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = v, REML=
  ↪ FALSE)

v.gesture <- lmer(targetAdvantage ~ gesture + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = v, REML=
  ↪ FALSE)

v.mannerFull <- lmer(targetAdvantage ~ speech*gesture + (1+
  ↪ speech+gesture|Participant) + (1|MediaName), data = v
  ↪ , REML=FALSE)

v.mannerSub <- lmer(targetAdvantage ~ speech + gesture +
  ↪ (1+speech+gesture|Participant) + (1|MediaName), data
  ↪ = v, REML=FALSE)

anova(v.null, v.speech)
Data: v
Models:
  v.null: targetAdvantage ~ 1 + (1 + speech + gesture |
  ↪ Participant) +
  v.null:      (1 | MediaName)
v.speech: targetAdvantage ~ speech + (1 + speech + gesture
  ↪ | Participant) +
  v.speech:      (1 | MediaName)
Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
v.null      9 72168 72243 -36075      72150
v.speech    10 72146 72230 -36063      72126 23.028      1
  ↪ 1.597e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

anova(v.null, v.gesture)
Data: v
Models:
  v.null: targetAdvantage ~ 1 + (1 + speech + gesture |
  ↪ Participant) + v.null: (1 | MediaName)

```

```

v.gesture: targetAdvantage ~ gesture + (1 + speech +
  ↳ gesture |Participant) + v.gesture: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
v.null      9 72168 72243 -36075      72150
v.gesture 10 72135 72218 -36057      72115 34.624      1
  ↳      4e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ '.' 1

anova(v.mannerSub, v.mannerFull)
Data: v
Models:
  v.mannerSub: targetAdvantage ~ speech + gesture + (1 +
    ↳ speech + gesture | v.mannerSub: Participant) +
    ↳ (1 | MediaName)
  v.mannerFull: targetAdvantage ~ speech * gesture + (1 +
    ↳ speech + gesture | v.mannerFull: Participant) + (1 |
    ↳ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
v.mannerSub  11 72097 72189 -36038      72075
v.mannerFull 12 72079 72179 -36027      72055 20.641      1
  ↳      5.54e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ '.' 1

lsmeans(v.mannerFull, pairwise~speech*gesture, adjust="
  ↳ tukey")
$lsmeans
speech gesture      lsmean          SE df  asymp.LCL  asymp.
  ↳ UCL
dotted absent -0.2979745 0.04540711 NA -0.3869708
  ↳ -0.2089782
wiggly absent  0.2608996 0.05378660 NA  0.1554798
  ↳ 0.3663194
dotted present 0.2827063 0.04721197 NA  0.1901725
  ↳ 0.3752400
wiggly present 0.4614245 0.05513051 NA  0.3533707
  ↳ 0.5694783

Confidence level used: 0.95

$contrasts
contrast              estimate          SE df z
  ↳ .ratio p.value
dotted,absent - wiggly,absent -0.55887404 0.06593304 NA
  ↳ -8.476 <.0001
dotted,absent - dotted,present -0.58068076 0.06261544 NA
  ↳ -9.274 <.0001

```

```
dotted,absent - wiggly,present -0.75939895 0.07022140 NA
  ↪ -10.814 <.0001
wiggly,absent - dotted,present -0.02180672 0.07049708 NA
  ↪ -0.309 0.9897
wiggly,absent - wiggly,present -0.20052491 0.06262884 NA
  ↪ -3.202 0.0075
dotted,present - wiggly,present -0.17871819 0.06529412 NA
  ↪ -2.737 0.0315
```

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(v.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: targetAdvantage ~ speech \* gesture + (1 + speech +  
 ↪ gesture | Participant) + (1 | MediaName)  
 Data: v

```
AIC      BIC    logLik deviance df.resid
72078.7  72178.7 -36027.3  72054.7    30855
```

Scaled **residuals**:

```
  Min      1Q  Median      3Q      Max
-2.5606 -0.7634  0.2611  0.7681  2.4308
```

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.06166	0.2483	
Participant	(Intercept)	0.01365	0.1168	
speechwiggly	0.03526	0.1878	-0.15	
gesturepresent	0.02187	0.1479	-0.43	-0.01
Residual		0.59007	0.7682	

Number of obs: 30867, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t value
(Intercept)		-0.29797
speechwiggly		0.55887
gesturepresent		0.58068
speechwiggly:gesturepresent		-0.38016

Correlation of Fixed Effects:

```
(Intr) spchwg gstrpr
speechwggly -0.587
gesturprsnt -0.660 0.399
spchwggly:g 0.451 -0.622 -0.646
```

#6800-7050



```
w <- subset(df1, Timebin >= 6750 & Timebin <= 7050)

w.null <- lmer(targetAdvantage ~ 1 + (1+speech+gesture|
  ↳ Participant) + (1|MediaName), data = w, REML=FALSE)
w.speech <- lmer(targetAdvantage ~ speech + (1+speech+
  ↳ gesture|Participant) + (1|MediaName), data = w, REML=
  ↳ FALSE)
w.gesture <- lmer(targetAdvantage ~ gesture + (1+speech+
  ↳ gesture|Participant) + (1|MediaName), data = w, REML=
  ↳ FALSE)
w.mannerFull <- lmer(targetAdvantage ~ speech*gesture + (1+
  ↳ speech+gesture|Participant) + (1|MediaName), data = w
  ↳ , REML=FALSE)
w.mannerSub <- lmer(targetAdvantage ~ speech + gesture +
  ↳ (1+speech+gesture|Participant) + (1|MediaName), data
  ↳ = w, REML=FALSE)

anova(w.null, w.speech)
Data: w
Models:
  w.null: targetAdvantage ~ 1 + (1 + speech + gesture |
    ↳ Participant) + w.null: (1 | MediaName)
  w.speech: targetAdvantage ~ speech + (1 + speech + gesture
    ↳ | Participant) + w.speech: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
w.null    9 75264 75339 -37623    75246
w.speech 10 75252 75336 -37616    75232 13.418      1
  ↳ 0.0002493 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

anova(w.null, w.gesture)
Data: w
Models:
  w.null: targetAdvantage ~ 1 + (1 + speech + gesture |
    ↳ Participant) + w.null: (1 | MediaName)
  w.gesture: targetAdvantage ~ gesture + (1 + speech +
    ↳ gesture | Participant) + w.gesture: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
w.null    9 75264 75339 -37623    75246
w.gesture 10 75224 75308 -37602    75204 41.497      1
  ↳ 1.181e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

anova(w.mannerSub, w.mannerFull)
Data: w
Models:
```

```

w.mannerSub: targetAdvantage ~ speech + gesture + (1 +
  ↳ speech + gesture | w.mannerSub: Participant) + (1 |
  ↳ MediaName)
w.mannerFull: targetAdvantage ~ speech * gesture + (1 +
  ↳ speech + gesture | w.mannerFull: Participant) +
  ↳ (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
w.mannerSub  11 75203 75295 -37591    75181
w.mannerFull 12 75184 75284 -37580    75160 21.244      1
  ↳ 4.044e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ '.' 1

lsmeans(w.mannerFull, pairwise~speech*gesture, adjust="
  ↳ tukey")
$lsmeans
speech gesture      lsmean          SE df asymp.LCL  asymp.
  ↳ UCL
dotted absent -0.2361395 0.04756926 NA -0.3293736
  ↳ -0.1429055
wiggly absent  0.2365509 0.05226442 NA  0.1341145
  ↳ 0.3389873
dotted present 0.3660831 0.05071612 NA  0.2666813
  ↳ 0.4654849
wiggly present 0.4439484 0.05542790 NA  0.3353117
  ↳ 0.5525851

Confidence level used: 0.95

$contrasts
contrast                                estimate          SE df z
  ↳ .ratio p.value
dotted,absent - wiggly,absent -0.47269040 0.06739036 NA
  ↳ -7.014 <.0001
dotted,absent - dotted,present -0.60222261 0.06451060 NA
  ↳ -9.335 <.0001
dotted,absent - wiggly,present -0.68008794 0.07257312 NA
  ↳ -9.371 <.0001
wiggly,absent - dotted,present -0.12953222 0.07211913 NA
  ↳ -1.796 0.2751
wiggly,absent - wiggly,present -0.20739754 0.06454453 NA
  ↳ -3.213 0.0072
dotted,present - wiggly,present -0.07786533 0.06674212 NA
  ↳ -1.167 0.6480

P value adjustment: tukey method for comparing a family of
  ↳ 4 estimates

summary(w.mannerFull)

```

```

Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: targetAdvantage ~ speech * gesture + (1 + speech +
  ↳ gesture | Participant) + (1 | MediaName)
Data: w

AIC      BIC    logLik deviance df.resid
75184.2  75284.4 -37580.1  75160.2    31281

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.4964 -0.7818  0.3010  0.7710  2.3196

Random effects:
  Groups      Name      Variance Std.Dev. Corr
MediaName (Intercept)  0.06449  0.2539
Participant (Intercept) 0.01798  0.1341
speechwiggly 0.03661  0.1913  -0.40
gesturepresent 0.02502  0.1582  -0.30  0.02
Residual      0.63136  0.7946
Number of obs: 31293, groups: MediaName, 158; Participant,
  ↳ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)          -0.23614      0.04757  -4.964
speechwiggly          0.47269      0.06739   7.014
gesturepresent        0.60222      0.06451   9.335
speechwiggly:gesturepresent -0.39483      0.08280  -4.769

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.635
gesturprnt -0.628  0.403
spchwggly:g  0.440 -0.622 -0.641

#7100-7350
x <- subset(df1, Timebin >= 7050 & Timebin <= 7350)

x.null <- lmer(targetAdvantage ~ 1 + (1+gesture+speech|
  ↳ Participant) + (1|MediaName), data = x, REML=FALSE)

x.speech <- lmer(targetAdvantage ~ speech + (1+gesture+
  ↳ speech|Participant) + (1|MediaName), data = x, REML=
  ↳ FALSE)

x.gesture <- lmer(targetAdvantage ~ gesture + (1+gesture+
  ↳ speech|Participant) + (1|MediaName), data = x, REML=
  ↳ FALSE)

x.mannerFull <- lmer(targetAdvantage ~ speech*gesture + (1+

```

```
↪ gesture+speech|Participant) + (1|MediaName), data = x
↪ , REML=FALSE)
```

```
x.mannerSub <- lmer(targetAdvantage ~ speech + gesture +
↪ (1+gesture+speech|Participant) + (1|MediaName), data
↪ = x, REML=FALSE)
```

```
anova(x.null, x.speech)
```

```
Data: x
```

```
Models:
```

```
x.null: targetAdvantage ~ 1 + (1 + gesture + speech |
↪ Participant) + x.null: (1 | MediaName)
```

```
x.speech: targetAdvantage ~ speech + (1 + gesture + speech
↪ | Participant) + x.speech: (1 | MediaName)
```

Df	AIC	BIC	logLik	<b>deviance</b>	Chisq	Chi	Df	Pr(>Chisq)
x.null	9	75721	75796	-37851	75703			
x.speech	10	75722	75805	-37851	75702	1.1233		1

↪ 0.2892

```
anova(x.null, x.gesture)
```

```
Data: x
```

```
Models:
```

```
x.null: targetAdvantage ~ 1 + (1 + gesture + speech |
↪ Participant) + x.null: (1 | MediaName)
```

```
x.gesture: targetAdvantage ~ gesture + (1 + gesture +
↪ speech | Participant) + x.gesture: (1 | MediaName
↪ )
```

Df	AIC	BIC	logLik	<b>deviance</b>	Chisq	Chi	Df	Pr(>Chisq)
x.null	9	75721	75796	-37851	75703			
x.gesture	10	75691	75775	-37835	75671	31.893		1

↪ 1.629e-08 \*\*\*

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
↪ ' ' 1
```

```
anova(x.mannerSub, x.mannerFull)
```

```
Data: x
```

```
Models:
```

```
x.mannerSub: targetAdvantage ~ speech + gesture + (1 +
↪ gesture + speech | x.mannerSub: Participant) + (1 |
↪ MediaName)
```

```
x.mannerFull: targetAdvantage ~ speech * gesture + (1 +
↪ gesture + speech | x.mannerFull: Participant) + (1 |
↪ MediaName)
```

Df	AIC	BIC	logLik	<b>deviance</b>	Chisq	Chi	Df	Pr(>Chisq)
x.mannerSub	11	75690	75782	-37834	75668			
x.mannerFull	12	75687	75787	-37831	75663	5.5261		1

↪ 0.01873 \*

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

↪ ‘ ’ 1

```
lsmeans(x.mannerFull, pairwise~speech*gesture, adjust="
↪ tukey")
```

**\$lsmeans**

speech gesture	lsmean	SE	df	asympt.LCL	asympt.
----------------	--------	----	----	------------	---------

↪ UCL

dotted absent	0.09030682	0.04982472	NA	-0.007347841	
---------------	------------	------------	----	--------------	--

↪ 0.1879615

wiggly absent	0.27557606	0.05257555	NA	0.172529871	
---------------	------------	------------	----	-------------	--

↪ 0.3786223

dotted present	0.48602961	0.04910221	NA	0.389791051	
----------------	------------	------------	----	-------------	--

↪ 0.5822682

wiggly present	0.47825485	0.05515038	NA	0.370162086	
----------------	------------	------------	----	-------------	--

↪ 0.5863476

Confidence level used: 0.95

**\$contrasts**

contrast		estimate	SE	df
----------	--	----------	----	----

↪ z.ratio p.value

dotted,absent - wiggly,absent		-0.185269248	0.06760151	NA
-------------------------------	--	--------------	------------	----

↪ -2.741 0.0312

dotted,absent - dotted,present		-0.395722797	0.06303360	NA
--------------------------------	--	--------------	------------	----

↪ -6.278 <.0001

dotted,absent - wiggly,present		-0.387948037	0.07443702	NA
--------------------------------	--	--------------	------------	----

↪ -5.212 <.0001

wiggly,absent - dotted,present		-0.210453550	0.06957160	NA
--------------------------------	--	--------------	------------	----

↪ -3.025 0.0133

wiggly,absent - wiggly,present		-0.202678789	0.06305579	NA
--------------------------------	--	--------------	------------	----

↪ -3.214 0.0072

dotted,present - wiggly,present		0.007774761	0.06699704	NA
---------------------------------	--	-------------	------------	----

↪ 0.116 0.9994

P value adjustment: tukey method **for** comparing a **family** of

↪ 4 estimates

**summary**(x.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']

Formula: targetAdvantage ~ speech \* gesture + (1 + gesture

↪ + speech | Participant) + (1 | MediaName)

Data: x

AIC	BIC	logLik	deviance	df.resid
75686.8	75787.2	-37831.4	75662.8	31528

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-2.6370	-0.8345	0.3540	0.7339	2.1898

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.06226	0.2495	
Participant	(Intercept)	0.02764	0.1662	
gesturepresent	0.02245	0.1498	-0.47	
speechwiggly	0.04172	0.2043	-0.47	0.20
Residual		0.62957	0.7935	

Number of obs: 31540, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t	value
(Intercept)	0.09031	0.04982	1.812
speechwiggly	0.18527	0.06760	2.741
gesturepresent	0.39572	0.06303	6.278
speechwiggly:gesturepresent	-0.19304	0.08140	-2.371

## Correlation of Fixed Effects:

(Intr)	spchwg	gstrpr
speechwggly	-0.637	
gesturprsnt	-0.644	0.435
spchwggly:g	0.413	-0.609 -0.645

#7350-7650

y &lt;- subset(df1, Timebin &gt;= 7350 &amp; Timebin &lt;= 7650)

y.null <- lmer(targetAdvantage ~ 1 + (1+speech+gesture|  
 ↪ Participant) + (1|MediaName), data = y, REML=FALSE)y.speech <- lmer(targetAdvantage ~ speech + (1+speech+  
 ↪ gesture|Participant) + (1|MediaName), data = y, REML=  
 ↪ FALSE)y.gesture <- lmer(targetAdvantage ~ gesture + (1+speech+  
 ↪ gesture|Participant) + (1|MediaName), data = y, REML=  
 ↪ FALSE)y.mannerFull <- lmer(targetAdvantage ~ speech\*gesture + (1+  
 ↪ speech+gesture|Participant) + (1|MediaName), data = y  
 ↪ , REML=FALSE)y.mannerSub <- lmer(targetAdvantage ~ speech + gesture +  
 ↪ (1+speech+gesture|Participant) + (1|MediaName), data  
 ↪ = y, REML=FALSE)

anova(y.null, y.speech)

Data: y

Models:

y.null: targetAdvantage ~ 1 + (1 + speech + gesture |  
 ↪ Participant) + y.null: (1 | MediaName)y.speech: targetAdvantage ~ speech + (1 + speech + gesture  
 ↪ | Participant) + y.speech: (1 | MediaName)

Df AIC BIC logLik deviance Chisq Chi Df Pr(&gt;Chisq)

y.null 9 68916 68991 -34449 68898

```

y.speech 10 68917 69000 -34448      68897 0.7896      1
  ↳ 0.3742

anova(y.null, y.gesture)
Data: y
Models:
  y.null: targetAdvantage ~ 1 + (1 + speech + gesture |
    ↳ Participant) +
  y.null:      (1 | MediaName)
y.gesture: targetAdvantage ~ gesture + (1 + speech +
  ↳ gesture | Participant) +
y.gesture:      (1 | MediaName)
Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
y.null      9 68916 68991 -34449      68898
y.gesture 10 68907 68990 -34443      68887 11.143      1
  ↳ 0.0008433 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

anova(y.mannerSub, y.mannerFull)
Data: y
Models:
  y.mannerSub: targetAdvantage ~ speech + gesture + (1 +
    ↳ speech + gesture | y.mannerSub: Participant) +
    ↳ (1 | MediaName)
y.mannerFull: targetAdvantage ~ speech * gesture + (1 +
  ↳ speech + gesture | y.mannerFull: Participant) + (1 |
  ↳ MediaName)
Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
y.mannerSub 11 68908 69000 -34443      68886
y.mannerFull 12 68910 69010 -34443      68886 0.2732      1
  ↳ 0.6012

lsmeans(y.mannerFull, pairwise~speech*gesture, adjust="
  ↳ tukey")
$lsmeans
speech gesture    lsmean          SE df asymp.LCL asymp.UCL
dotted absent  0.4349733 0.04607668 NA 0.3446647 0.5252820
wiggly absent  0.4294189 0.04398211 NA 0.3432155 0.5156222
dotted present 0.5845658 0.04372734 NA 0.4988617 0.6702698
wiggly present 0.5466853 0.04637790 NA 0.4557863 0.6375843

Confidence level used: 0.95

$contrasts
contrast                estimate          SE df
  ↳ z.ratio p.value
dotted,absent - wiggly,absent  0.005554458 0.05307316 NA

```



```

↪ 0.105 0.9996
dotted,absent - dotted,present -0.149592406 0.05063115 NA
↪ -2.955 0.0165
dotted,absent - wiggly,present -0.111711966 0.06225248 NA
↪ -1.794 0.2759
wiggly,absent - dotted,present -0.155146864 0.05493534 NA
↪ -2.824 0.0245
wiggly,absent - wiggly,present -0.117266423 0.05064559 NA
↪ -2.315 0.0945
dotted,present - wiggly,present 0.037880441 0.05263275 NA
↪ 0.720 0.8893

```

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(y.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: targetAdvantage ~ speech \* gesture + (1 + speech +  
 ↪ gesture | Participant) + (1 | MediaName)  
 Data: y

AIC	BIC	logLik	<b>deviance</b>	<b>df.resid</b>
68910.0	69010.2	-34443.0	68886.0	31180

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-2.8558	-0.5126	0.3658	0.6643	1.9531

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.03509	0.1873	
Participant	(Intercept)	0.03990	0.1998	
speechwiggly	0.03021	0.1738	-0.53	
gesturepresent	0.02225	0.1492	-0.48	0.29
Residual		0.52094	0.7218	

Number of obs: 31192, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	<b>t</b>	value
(Intercept)		0.434973	0.046077 9.440
speechwiggly		-0.005554	0.053073 -0.105
gesturepresent		0.149592	0.050631 2.955
speechwiggly:gesturepresent		-0.032326	0.061824 -0.523

Correlation of Fixed Effects:

(Intr)	spchwg	gstrpr
speechwggly	-0.614	
gesturprsnt	-0.595	0.440
spchwgly:g	0.339	-0.590 -0.610



```
#7700-7950
```

```
z <- subset(df1, Timebin >= 7650 & Timebin <= 7950)
```

```
z.null <- lmer(targetAdvantage ~ 1 + (1+speech+gesture|
  ↳ Participant) + (1|MediaName), data = z, REML=FALSE)
```

```
z.speech <- lmer(targetAdvantage ~ speech + (1+speech+
  ↳ gesture|Participant) + (1|MediaName), data = z, REML=
  ↳ FALSE)
```

```
z.gesture <- lmer(targetAdvantage ~ gesture + (1+speech+
  ↳ gesture|Participant) + (1|MediaName), data = z, REML=
  ↳ FALSE)
```

```
z.mannerFull <- lmer(targetAdvantage ~ speech*gesture + (1+
  ↳ speech+gesture|Participant) + (1|MediaName), data = z
  ↳ , REML=FALSE)
```

```
z.mannerSub <- lmer(targetAdvantage ~ speech + gesture +
  ↳ (1+speech+gesture|Participant) + (1|MediaName), data
  ↳ = z, REML=FALSE)
```

```
anova(z.null, z.gesture)
```

```
Data: z
```

```
Models:
```

```
z.null: targetAdvantage ~ 1 + (1 + speech + gesture |
  ↳ Participant) + z.null: (1 | MediaName)
```

```
z.gesture: targetAdvantage ~ gesture + (1 + speech +
  ↳ gesture | Participant) + z.gesture: (1 | MediaName)
```

```
Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
z.null 9 59148 59223 -29565 59130
```

```
z.gesture 10 59150 59233 -29565 59130 0.436 1
```

```
↳ 0.5091
```

```
anova(z.null, z.speech)
```

```
Data: z
```

```
Models:
```

```
z.null: targetAdvantage ~ 1 + (1 + speech + gesture |
  ↳ Participant) + z.null: (1 | MediaName)
```

```
z.speech: targetAdvantage ~ speech + (1 + speech + gesture
  ↳ | Participant) + z.speech: (1 | MediaName)
```

```
Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
z.null 9 59148 59223 -29565 59130
```

```
z.speech 10 59145 59228 -29562 59125 5.3348 1
```

```
↳ 0.0209 *
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
↳ ' ' 1
```

```
anova(z.mannerSub, z.mannerFull)
```

```
Data: z
```

```
Models:
```

```

z.mannerSub: targetAdvantage ~ speech + gesture + (1 +
  ↳ speech + gesture | z.mannerSub: Participant) + (1 |
  ↳ MediaName)
z.mannerFull: targetAdvantage ~ speech * gesture + (1 +
  ↳ speech + gesture | z.mannerFull: Participant) + (1 |
  ↳ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
z.mannerSub  11 59146 59237 -29562    59124
z.mannerFull 12 59142 59242 -29559    59118 5.8089      1
  ↳    0.01594 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳  ' ' 1

lsmeans(z.mannerFull, pairwise~speech*gesture, adjust="
  ↳ tukey")
$lsmeans
speech gesture    lsmean          SE df asymp.LCL asymp.UCL
dotted absent  0.6509413 0.03997294 NA 0.5725957 0.7292868
wiggly absent  0.5008525 0.04302607 NA 0.4165230 0.5851821
dotted present 0.5666118 0.04205278 NA 0.4841899 0.6490337
wiggly present 0.5204762 0.04863863 NA 0.4251463 0.6158062

Confidence level used: 0.95

$contrasts
contrast                                estimate          SE df z
  ↳ .ratio p.value
dotted,absent - wiggly,absent           0.15008872 0.04381988 NA
  ↳    3.425  0.0034
dotted,absent - dotted,present          0.08432947 0.03644843 NA
  ↳    2.314  0.0949
dotted,absent - wiggly,present          0.13046504 0.05167420 NA
  ↳    2.525  0.0562
wiggly,absent - dotted,present          -0.06575925 0.04449435 NA
  ↳   -1.478  0.4509
wiggly,absent - wiggly,present          -0.01962369 0.03644799 NA
  ↳   -0.538  0.9497
dotted,present - wiggly,present          0.04613556 0.04357399 NA
  ↳    1.059  0.7146

P value adjustment: tukey method for comparing a family of
  ↳ 4 estimates

summary(z.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: targetAdvantage ~ speech * gesture + (1 + speech +
  ↳ gesture | Participant) + (1 | MediaName)
Data: z

```

```

AIC      BIC    logLik deviance df.resid
59142.0  59241.8 -29559.0  59118.0    30313

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.3134 -0.3790  0.3140  0.6422  1.7570

Random effects:
  Groups      Name              Variance Std.Dev. Corr
MediaName (Intercept)    0.01592  0.1262
Participant (Intercept) 0.03935  0.1984
speechwiggly 0.03430  0.1852  -0.35
gesturepresent 0.01411  0.1188  -0.17  0.28
Residual                    0.40255  0.6345
Number of obs: 30325, groups:  MediaName, 158; Participant,
  ↪ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)                0.65094    0.03997  16.285
speechwiggly                -0.15009    0.04382  -3.425
gesturepresent              -0.08433    0.03645  -2.314
speechwiggly:gesturepresent  0.10395    0.04273   2.433

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.476
gesturprnt -0.397  0.397
spchwggly:g  0.270 -0.493 -0.586

##Character Advantage

#3 create a df only including character prop

g <- e[e$AOI == "AOI.Character.Hit",]

#4. create a column containing the average proportion of
  ↪ looks to character per item per participant

h <- ddply(g,.(MediaName, Participant, Timebin),transform,
  ↪ CharacterProp=mean(prop))

#5.create a new df containing only those columns of
  ↪ interest including the new avDisProp column

j <- h[, c("Participant", "MediaName", "condition", "
  ↪ Timebin", "gesture", "speech", "CharacterProp")]

df2 <- merge(i, j, all=T)

```

```

df2$characterAdvantage <- df2$Character - df2$avDisProp

write.csv(df2, "characterAdvantage.csv")
df2 <- read.csv("characterAdvantage.csv", header = T)

#Plotting Character advantage
scatterCa <- ggplot(df2, aes(Timebin, characterAdvantage,
  ↪ colour = condition))

scatterCa + stat_summary(fun.y = mean, geom = "point") +
  ↪ stat_summary(fun.y = mean, geom = "line", aes (group
  ↪ = condition)) + labs(x = "Time(ms)", y = "Character_
  ↪ Advantage", colour = "condition") + ggtitle("
  ↪ Character_Advantage") + annotate("rect", xmin = 5270,
  ↪ xmax = 5540, ymin = -0.50, ymax = 0.9, alpha = .2)+
  ↪ annotate("rect", xmin = 6950, xmax = 7500, ymin =
  ↪ -0.5, ymax = 0.9, alpha = .2) + scale_color_manual("
  ↪ Condition", breaks=c("condition1","condition2", "
  ↪ condition3", "condition4"), labels=c("+Speech;_+
  ↪ Gesture", "+Speech;_+Gesture", "-Speech;_+Gesture", "
  ↪ -Speech;_+Gesture"), values = c("red", "blue", "green
  ↪ ", "darkgreen")) + theme(legend.text=element_text(
  ↪ size=12), axis.text=element_text(size=12),axis.title=
  ↪ element_text(size=14,face="bold"),legend.title=
  ↪ element_text(size=14),plot.title=element_text(size
  ↪ =16)) + annotate("text", x = 5100, y = 0.1, label = "
  ↪ mannerDOP") + annotate("text", x = 8000, y = 0.50,
  ↪ label = "groundDOP")

#5250-5550
rc<- subset(df2, Timebin >= 5250 & Timebin <= 5550)

rc.null <- lmer(characterAdvantage ~ 1 + (1+speech+gesture|
  ↪ Participant) + (1|MediaName), data = rc, REML=FALSE)

rc.speech <- lmer(characterAdvantage ~ speech + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = rc, REML
  ↪ =FALSE)

rc.gesture <- lmer(characterAdvantage ~ gesture + (1+speech
  ↪ +gesture|Participant) + (1|MediaName), data = rc,
  ↪ REML=FALSE)

rc.mannerFull <- lmer(characterAdvantage ~ speech*gesture +
  ↪ (1+speech+gesture|Participant) + (1|MediaName), data
  ↪ = rc, REML=FALSE)

rc.mannerSub <- lmer(characterAdvantage ~ speech + gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = rc, REML=FALSE)

```

```

anova(rc.null, rc.gesture)
Data: rc
Models:
  rc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) + rc.null:(1 | MediaName)
  rc.gesture: characterAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + rc.gesture: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
rc.null      9 43177 43252 -21580    43159
rc.gesture 10 43179 43262 -21579    43159 0.587      1
    ↪ 0.4436

anova(rc.null, rc.speech)
Data: rc
Models:
  rc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) + rc.null:      (1 | MediaName)
  rc.speech: characterAdvantage ~ speech + (1 + speech +
    ↪ gesture | Participant) + rc.speech:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
rc.null      9 43177 43252 -21580    43159
rc.speech 10 43179 43263 -21580    43159 0.0675     1
    ↪ 0.7951

anova(rc.mannerSub, rc.mannerFull)
Data: rc
Models:
  rc.mannerSub: characterAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | rc.mannerSub:      Participant)
    ↪ + (1 | MediaName)
  rc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
    ↪ speech + gesture | rc.mannerFull: Participant) + (1
    ↪ | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
rc.mannerSub 11 43181 43272 -21579    43159
rc.mannerFull 12 43181 43281 -21578    43157 1.8237     1
    ↪      0.1769

summary(rc.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: characterAdvantage ~ speech * gesture + (1 +
    ↪ speech + gesture |      Participant) + (1 | MediaName
    ↪ )
Data: rc

AIC      BIC   logLik deviance df.resid
43180.7  43280.9 -21578.4  43156.7    31056

```

```

Scaled residuals:
  Min      1Q  Median      3Q      Max
-4.2980  0.0016  0.1719  0.4026  2.3190

Random effects:
  Groups      Name              Variance Std.Dev. Corr
MediaName    (Intercept)         0.010031 0.10016
Participant  (Intercept)         0.028232 0.16802
speechwiggly 0.007559 0.08694 -0.25
gesturepresent 0.004549 0.06745  0.08  0.42
Residual                                0.230121 0.47971
Number of obs: 31068, groups: MediaName, 158; Participant,
  ↪ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)                0.83512    0.03319  25.164
speechwiggly                -0.03173    0.02825  -1.123
gesturepresent              -0.03965    0.02651  -1.496
speechwiggly:gesturepresent  0.04582    0.03372   1.359

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.422
gesturprsnt -0.299  0.479
spchwggly:g  0.257 -0.603 -0.635

lsmeans(rc.mannerFull, pairwise~speech*gesture, adjust="
  ↪ tukey")
$lsmeans
speech gesture    lsmean      SE df asymp.LCL asymp.UCL
dotted absent  0.8351168 0.03318760 NA 0.7700703 0.9001633
wiggly absent  0.8033899 0.03329006 NA 0.7381426 0.8686372
dotted present 0.7954664 0.03575393 NA 0.7253900 0.8655428
wiggly present 0.8095614 0.03781158 NA 0.7354521 0.8836707

Confidence level used: 0.95

$contrasts
contrast                                estimate      SE df
  ↪ z.ratio p.value
dotted,absent - wiggly,absent          0.031726901 0.02824579 NA
  ↪ 1.123 0.6752
dotted,absent - dotted,present         0.039650417 0.02651118 NA
  ↪ 1.496 0.4401
dotted,absent - wiggly,present         0.025555408 0.03272185 NA
  ↪ 0.781 0.8631
wiggly,absent - dotted,present         0.007923517 0.02797784 NA
  ↪ 0.283 0.9921
wiggly,absent - wiggly,present        -0.006171493 0.02653856 NA

```

```

    ↪ -0.233  0.9956
dotted,present - wiggly,present -0.014095010 0.02802838 NA
    ↪ -0.503  0.9584

P value adjustment: tukey method for comparing a family of
    ↪ 4 estimates

#5550 - 5850
sc<- subset(df2, Timebin >= 5550 & Timebin <= 5850)

sc.null <- lmer(characterAdvantage ~ 1 + (1+gesture+speech|
    ↪ Participant) + (1|MediaName), data = sc, REML=FALSE)

sc.speech <- lmer(characterAdvantage ~ speech + (1+gesture+
    ↪ speech|Participant) + (1|MediaName), data = sc, REML=
    ↪ FALSE)

sc.gesture <- lmer(characterAdvantage ~ gesture + (1+
    ↪ gesture+speech|Participant) + (1|MediaName), data =
    ↪ sc, REML=FALSE)

sc.mannerFull <- lmer(characterAdvantage ~ speech*gesture +
    ↪ (1+gesture+speech|Participant) + (1|MediaName), data
    ↪ = sc, REML=FALSE)

sc.mannerSub <- lmer(characterAdvantage ~ speech + gesture
    ↪ + (1+gesture+speech|Participant) + (1|MediaName),
    ↪ data = sc, REML=FALSE)

anova(sc.null, sc.gesture)
Data: sc
Models:
  sc.null: characterAdvantage ~ 1 + (1 + gesture + speech |
    ↪ Participant) + sc.null: (1 | MediaName)
  sc.gesture: characterAdvantage ~ gesture + (1 + gesture +
    ↪ speech | Participant) + sc.gesture: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
sc.null    9 60137 60212 -30059    60119
sc.gesture 10 60138 60222 -30059    60118 0.1773    1
    ↪    0.6737

anova(sc.null, sc.speech)
Data: sc
Models:
  sc.null: characterAdvantage ~ 1 + (1 + gesture + speech |
    ↪ Participant) + sc.null: (1 | MediaName)
  sc.speech: characterAdvantage ~ speech + (1 + gesture +
    ↪ speech | Participant) + sc.speech: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
sc.null    9 60137 60212 -30059    60119

```



```

sc.speech 10 60131 60215 -30056      60111 7.3273      1
  ↪ 0.006791 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ '.' 1

anova(sc.mannerSub, sc.mannerFull)
Data: sc
Models:
  sc.mannerSub: characterAdvantage ~ speech + gesture + (1
  ↪ + gesture + speech | sc.mannerSub: Participant) +
  ↪ (1 | MediaName)
  sc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
  ↪ gesture + speech | sc.mannerFull: Participant) + (1
  ↪ | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
sc.mannerSub  11 60133 60225 -30056      60111
sc.mannerFull 12 60135 60235 -30056      60111 0.0239      1
  ↪      0.8772

lsmeans(sc.mannerFull, pairwise~speech*gesture, adjust="
  ↪ tukey")
$lsmeans
speech gesture    lsmean          SE df asymp.LCL asymp.UCL
dotted absent  0.6558371 0.04771257 NA 0.5623222 0.7493521
wiggly absent  0.5498476 0.05968920 NA 0.4328590 0.6668363
dotted present 0.6371790 0.05125367 NA 0.5367236 0.7376343
wiggly present 0.5401741 0.06256703 NA 0.4175450 0.6628033

Confidence level used: 0.95

$contrasts
contrast                estimate          SE df
  ↪ z.ratio p.value
dotted,absent - wiggly,absent    0.105989491 0.04736746 NA
  ↪ 2.238 0.1132
dotted,absent - dotted,present    0.018658149 0.04388896 NA
  ↪ 0.425 0.9742
dotted,absent - wiggly,present    0.115662997 0.04960519 NA
  ↪ 2.332 0.0909
wiggly,absent - dotted,present   -0.087331342 0.04958735 NA
  ↪ -1.761 0.2922
wiggly,absent - wiggly,present    0.009673505 0.04390541 NA
  ↪ 0.220 0.9962
dotted,present - wiggly,present    0.097004848 0.04693194 NA
  ↪ 2.067 0.1640

P value adjustment: tukey method for comparing a family of
  ↪ 4 estimates

```



```

summary(sc.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: characterAdvantage ~ speech * gesture + (1 +
  ↳ gesture + speech | Participant) + (1 | MediaName
  ↳ )
Data: sc

AIC      BIC    logLik deviance df.resid
60135.0  60235.3 -30055.5  60111.0    31445

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.3437 -0.1535  0.2480  0.6075  2.6428

Random effects:
  Groups      Name          Variance Std.Dev. Corr
MediaName (Intercept)    0.031408 0.17722
Participant (Intercept)  0.049381 0.22222
gesturepresent 0.007932 0.08906  0.13
speechwiggly  0.018217 0.13497  0.45 0.00
Residual                                0.386529 0.62171
Number of obs: 31457, groups: MediaName, 158; Participant,
  ↳ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)                0.655837  0.047713  13.746
speechwiggly                -0.105989  0.047367  -2.238
gesturepresent              -0.018658  0.043889  -0.425
speechwiggly:gesturepresent  0.008985  0.058143   0.155

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.212
gesturprsnt -0.376  0.412
spchwggly:g  0.308 -0.621 -0.662

#5900-6150
tc <- subset(df2, Timebin >= 5850 & Timebin <= 6150)

tc.null <- lmer(characterAdvantage ~ 1 + (1+gesture+speech|
  ↳ Participant) + (1|MediaName), data = tc, REML=FALSE)

tc.speech <- lmer(characterAdvantage ~ speech + (1+gesture+
  ↳ speech|Participant) + (1|MediaName), data = tc, REML=
  ↳ FALSE)

tc.gesture <- lmer(characterAdvantage ~ gesture + (1+
  ↳ gesture+speech|Participant) + (1|MediaName), data =
  ↳ tc, REML=FALSE)

```

```
tc.mannerFull <- lmer(characterAdvantage ~ speech*gesture +
  ↪ (1+gesture+speech|Participant) + (1|MediaName), data
  ↪ = tc, REML=FALSE)
```

```
tc.mannerSub <- lmer(characterAdvantage ~ speech + gesture
  ↪ + (1+gesture+speech|Participant) + (1|MediaName),
  ↪ data = tc, REML=FALSE)
```

```
anova(tc.null, tc.gesture)
```

```
Data: tc
```

```
Models:
```

```
tc.null: characterAdvantage ~ 1 + (1 + gesture + speech |
  ↪ Participant) + tc.null: (1 | MediaName)
```

```
tc.gesture: characterAdvantage ~ gesture + (1 + gesture +
  ↪ speech | Participant) + tc.gesture: (1 | MediaName)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
tc.null	9	66600	66675	-33291	66582			
tc.gesture	10	66597	66680	-33288	66577	5.4501	1	↪ 0.01957 *

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1
```

```
anova(tc.null, tc.speech)
```

```
Data: tc
```

```
Models:
```

```
tc.null: characterAdvantage ~ 1 + (1 + gesture + speech |
  ↪ Participant) + tc.null: (1 | MediaName)
```

```
tc.speech: characterAdvantage ~ speech + (1 + gesture +
  ↪ speech | Participant) + tc.speech: (1 | MediaName)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
tc.null	9	66600	66675	-33291	66582			
tc.speech	10	66590	66673	-33285	66570	12.589	1	↪ 0.000388 ***

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1
```

```
anova(tc.mannerSub, tc.mannerFull)
```

```
Data: tc
```

```
Models:
```

```
tc.mannerSub: characterAdvantage ~ speech + gesture + (1
  ↪ + gesture + speech | tc.mannerSub: Participant)
  ↪ + (1 | MediaName)
```

```
tc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
  ↪ gesture + speech | tc.mannerFull: Participant) +
  ↪ (1 | MediaName)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
tc.mannerSub	11	66587	66679	-33282	66565			

```

tc.mannerFull 12 66585 66685 -33280      66561 4.4752      1
  ↳ 0.03439 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ '.' 1

lsmeans(tc.mannerFull, pairwise~speech*gesture, adjust="
  ↳ tukey")
$lsmeans
speech gesture    lsmean          SE df  asymp.LCL asymp.UCL
dotted absent  0.2980853 0.06060167 NA 0.17930823 0.4168624
wiggly absent  0.2054944 0.06320337 NA 0.08161806 0.3293707
dotted present 0.4623423 0.05852705 NA 0.34763139 0.5770532
wiggly present 0.2282260 0.06374857 NA 0.10328109 0.3531709

Confidence level used: 0.95

$contrasts
contrast                estimate          SE df z
  ↳ .ratio p.value
dotted,absent - wiggly,absent    0.09259093 0.05721570 NA
  ↳ 1.618 0.3682
dotted,absent - dotted,present  -0.16425698 0.05386118 NA
  ↳ -3.050 0.0123
dotted,absent - wiggly,present    0.06985932 0.06522690 NA
  ↳ 1.071 0.7072
wiggly,absent - dotted,present   -0.25684790 0.06018790 NA
  ↳ -4.267 0.0001
wiggly,absent - wiggly,present   -0.02273161 0.05384200 NA
  ↳ -0.422 0.9747
dotted,present - wiggly,present   0.23411630 0.05670633 NA
  ↳ 4.129 0.0002

P value adjustment: tukey method for comparing a family of
  ↳ 4 estimates

summary(tc.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: characterAdvantage ~ speech * gesture + (1 +
  ↳ gesture + speech | Participant) + (1 | MediaName)
Data: tc

AIC      BIC    logLik deviance df.resid
66584.5  66684.7 -33280.3  66560.5    31221

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.9486 -0.6692  0.2223  0.7493  2.7892

Random effects:

```

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.04112	0.2028	
Participant	(Intercept)	0.08875	0.2979	
gesturepresent	0.02364	0.1537	-0.34	
speechwiggly	0.03562	0.1887	-0.22	0.19
Residual		0.48078	0.6934	

Number of obs: 31233, groups: MediaName, 158; Participant, 35

Fixed **effects**:

Estimate	Std. Error	t	value
(Intercept)		0.29809	0.06060 4.919
speechwiggly		-0.09259	0.05722 -1.618
gesturepresent		0.16426	0.05386 3.050
speechwiggly:gesturepresent		-0.14153	0.06642 -2.131

Correlation of Fixed Effects:

(Intr)	spchwg	gstrpr
speechwggly	-0.426	
gesturprsnt	-0.482	0.414
spchwggly:g	0.278	-0.588 -0.617

#6150-6450

```
uc <- subset(df2, Timebin >= 6150 & Timebin <= 6450)
```

```
uc.null <- lmer(characterAdvantage ~ 1 + (1+speech+gesture|
  ↪ Participant) + (1|MediaName), data = uc, REML=FALSE)
```

```
uc.speech <- lmer(characterAdvantage ~ speech + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = uc, REML
  ↪ =FALSE)
```

```
uc.gesture <- lmer(characterAdvantage ~ gesture + (1+speech
  ↪ +gesture|Participant) + (1|MediaName), data = uc,
  ↪ REML=FALSE)
```

```
uc.mannerFull <- lmer(characterAdvantage ~ speech*gesture +
  ↪ (1+speech+gesture|Participant) + (1|MediaName), data
  ↪ = uc, REML=FALSE)
```

```
uc.mannerSub <- lmer(characterAdvantage ~ speech + gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = uc, REML=FALSE)
```

```
anova(uc.null, uc.speech)
```

Data: uc

Models:

```
uc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
  ↪ Participant) + uc.null: (1 | MediaName)
uc.speech: characterAdvantage ~ speech + (1 + speech +
```

```

    ↪ gesture | Participant) + uc.speech: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
uc.null      9 64437 64512 -32210      64419
uc.speech 10 64437 64521 -32209      64417 1.9374      1
    ↪ 0.1639

anova(uc.null, uc.gesture)
Data: uc
Models:
  uc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) + uc.null: (1 | MediaName)
  uc.gesture: characterAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + uc.gesture: (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
uc.null      9 64437 64512 -32210      64419
uc.gesture 10 64412 64496 -32196      64392 27.053      1
    ↪ 1.979e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

anova(uc.mannerSub, uc.mannerFull)
Data: uc
Models:
  uc.mannerSub: characterAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | uc.mannerSub: Participant)
    ↪ + (1 | MediaName)
  uc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
    ↪ speech + gesture | uc.mannerFull: Participant) +
    ↪ (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
uc.mannerSub 11 64414 64505 -32196      64392
uc.mannerFull 12 64409 64509 -32193      64385 6.6621      1
    ↪ 0.009849 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

lsmeans(uc.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL  asymp
    ↪ .UCL
dotted absent -0.07979178 0.06057102 NA -0.19850880
    ↪ 0.03892523
wiggly absent -0.03493881 0.04693973 NA -0.12693899
    ↪ 0.05706137
dotted present 0.20540653 0.05494170 NA 0.09772278
    ↪ 0.31309028

```

wiggly present 0.09979476 0.04492400 NA 0.01174535  
 ↪ 0.18784418

Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z
↪ .ratio p.value				
dotted,absent - wiggly,absent	-0.04485297	0.05078938	NA	
↪ -0.883 0.8136				
dotted,absent - dotted,present	-0.28519831	0.04794996	NA	
↪ -5.948 <.0001				
dotted,absent - wiggly,present	-0.17958655	0.06049480	NA	
↪ -2.969 0.0158				
wiggly,absent - dotted,present	-0.24034534	0.05223288	NA	
↪ -4.601 <.0001				
wiggly,absent - wiggly,present	-0.13473358	0.04795530	NA	
↪ -2.810 0.0256				
dotted,present - wiggly,present	0.10561176	0.05036159	NA	
↪ 2.097 0.1540				

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(uc.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: characterAdvantage ~ speech \* gesture + (1 +  
 ↪ speech + gesture | Participant) + (1 | MediaName)  
 Data: uc

AIC	BIC	logLik	deviance	df.resid
64409.0	64509.1	-32192.5	64385.0	31003

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-2.9709	-0.7151	0.0264	0.7850	2.7547

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.03052	0.1747	
Participant	(Intercept)	0.09838	0.3137	
speechwiggly	0.03073	0.1753	-0.75	
gesturepresent	0.02170	0.1473	-0.47	0.32
Residual		0.45584	0.6752	

Number of obs: 31015, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t value
(Intercept)	-0.07979	0.06057 -1.317

```
speechwiggly          0.04485    0.05079    0.883
gesturepresent        0.28520    0.04795    5.948
speechwiggly:gesturepresent -0.15046    0.05768   -2.609
```

Correlation of Fixed Effects:

```
(Intr) spchwg gstrpr
speechwggly -0.657
gesturprsnt -0.508  0.442
spchwggly:g  0.241 -0.575 -0.601
```

#6450-6750

```
vc <- subset(df2, Timebin >= 6450 & Timebin <= 6750)
```

```
vc.null <- lmer(characterAdvantage ~ 1 + (1+speech+gesture|
  ↳ Participant) + (1|MediaName), data = vc, REML=FALSE)
```

```
vc.speech <- lmer(characterAdvantage ~ speech + (1+speech+
  ↳ gesture|Participant) + (1|MediaName), data = vc, REML
  ↳ =FALSE)
```

```
vc.gesture <- lmer(characterAdvantage ~ gesture + (1+speech
  ↳ +gesture|Participant) + (1|MediaName), data = vc,
  ↳ REML=FALSE)
```

```
vc.mannerFull <- lmer(characterAdvantage ~ speech*gesture +
  ↳ (1+speech+gesture|Participant) + (1|MediaName), data
  ↳ = vc, REML=FALSE)
```

```
vc.mannerSub <- lmer(characterAdvantage ~ speech + gesture
  ↳ + (1+speech+gesture|Participant) + (1|MediaName),
  ↳ data = vc, REML=FALSE)
```

```
anova(vc.null, vc.speech)
```

Data: vc

Models:

```
vc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
  ↳ Participant) +
vc.null:      (1 | MediaName)
```

```
vc.speech: characterAdvantage ~ speech + (1 + speech +
  ↳ gesture | Participant) +
```

```
vc.speech:      (1 | MediaName)
```

```
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
vc.null      9 58836 58911 -29409    58818
```

```
vc.speech  10 58837 58920 -29408    58817 1.0224      1
```

```
↳      0.312
```

```
anova(vc.null, vc.gesture)
```

Data: vc

Models:

```
vc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
```

```

    ↪ Participant) + vc.null: (1 | MediaName)
vc.gesture: characterAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + vc.gesture: (1 | MediaName)
Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
vc.null      9 58836 58911 -29409      58818
vc.gesture 10 58792 58876 -29386      58772 45.5      1
    ↪ 1.526e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

anova(vc.mannerSub, vc.mannerFull)
Data: vc
Models:
  vc.mannerSub: characterAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | vc.mannerSub: Participant) +
    ↪ (1 | MediaName)
  vc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
    ↪ speech + gesture | vc.mannerFull: Participant) + (1
    ↪ | MediaName)
Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
vc.mannerSub 11 58787 58878 -29382      58765
vc.mannerFull 12 58784 58884 -29380      58760 4.742      1
    ↪ 0.02944 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

lsmeans(vc.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL
    ↪ asymp.UCL
dotted absent -0.35169934 0.05192649 NA -0.45347340
    ↪ -0.24992529
wiggly absent -0.19153320 0.04203808 NA -0.27392631
    ↪ -0.10914008
dotted present -0.03108267 0.04705073 NA -0.12330040
    ↪ 0.06113506
wiggly present 0.01687647 0.03903792 NA -0.05963645
    ↪ 0.09338939

Confidence level used: 0.95

$contrasts
contrast              estimate      SE df z
    ↪ .ratio p.value
dotted,absent - wiggly,absent -0.16016614 0.04489814 NA
    ↪ -3.567 0.0020
dotted,absent - dotted,present -0.32061667 0.04084325 NA

```



```

    ↪ -7.850 <.0001
dotted,absent - wiggly,present -0.36857581 0.05099835 NA
    ↪ -7.227 <.0001
wiggly,absent - dotted,present -0.16045053 0.04604416 NA
    ↪ -3.485 0.0028
wiggly,absent - wiggly,present -0.20840967 0.04085801 NA
    ↪ -5.101 <.0001
dotted,present - wiggly,present -0.04795914 0.04452589 NA
    ↪ -1.077 0.7035

```

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(vc.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']

Formula: characterAdvantage ~ speech \* gesture + (1 +

↪ speech + gesture | Participant) + (1 | MediaName)

Data: vc

```

AIC      BIC    logLik deviance df.resid
58784.0  58884.1 -29380.0  58760.0    30855

```

Scaled **residuals**:

```

  Min      1Q  Median      3Q      Max
-2.5852 -0.7073  0.0108  0.5689  3.0578

```

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.02384	0.1544	
Participant	(Intercept)	0.07074	0.2660	
speechwiggly	0.02370	0.1539	-0.69	
gesturepresent	0.01223	0.1106	-0.48	0.25
Residual		0.38401	0.6197	

Number of obs: 30867, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

	Estimate	Std. Error	t value
(Intercept)	-0.35170	0.05193	-6.773
speechwiggly	0.16017	0.04490	3.567
gesturepresent	0.32062	0.04084	7.850
speechwiggly:gesturepresent	-0.11221	0.05114	-2.194

Correlation of Fixed Effects:

```

(Intr) spchwg gstrpr
speechwggly -0.632
gesturprsnt -0.507  0.426
spchwggly:g  0.249 -0.577 -0.626

```

#6800-7050

```

wc <- subset(df2, Timebin >= 6750 & Timebin <= 7050)

wc.null <- lmer(characterAdvantage ~ 1 + (1+speech+gesture|
  ↪ Participant) + (1|MediaName), data = wc, REML=FALSE)
wc.speech <- lmer(characterAdvantage ~ speech + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = wc, REML
  ↪ =FALSE)
wc.gesture <- lmer(characterAdvantage ~ gesture + (1+speech
  ↪ +gesture|Participant) + (1|MediaName), data = wc,
  ↪ REML=FALSE)
wc.mannerFull <- lmer(characterAdvantage ~ speech*gesture +
  ↪ (1+speech+gesture|Participant) + (1|MediaName), data
  ↪ = wc, REML=FALSE)
wc.mannerSub <- lmer(characterAdvantage ~ speech + gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = wc, REML=FALSE)

anova(wc.null, wc.speech)
Data: wc
Models:
  wc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) + wc.null: (1 | MediaName)
  wc.speech: characterAdvantage ~ speech + (1 + speech +
    ↪ gesture | Participant) + wc.speech: (1 | MediaName)
Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
wc.null      9 55311 55386 -27646      55293
wc.speech    10 55297 55381 -27639      55277 15.703      1
  ↪ 7.412e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

anova(wc.null, wc.gesture)
Data: wc
Models:
  wc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) + wc.null: (1 | MediaName)
  wc.gesture: characterAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + wc.gesture: (1 | MediaName)
Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
wc.null      9 55311 55386 -27646      55293
wc.gesture   10 55278 55361 -27629      55258 34.939      1
  ↪ 3.403e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

anova(wc.mannerSub, wc.mannerFull)
Data: wc
Models:

```

```

wc.mannerSub: characterAdvantage ~ speech + gesture + (1
  ↳ + speech + gesture | wc.mannerSub: Participant) +
  ↳ (1 | MediaName)
wc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
  ↳ speech + gesture | wc.mannerFull: Participant) + (1
  ↳ | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
wc.mannerSub  11 55250 55342 -27614      55228
wc.mannerFull 12 55245 55346 -27611      55221 6.583      1
  ↳      0.0103 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ ' ' 1

lsmeans(wc.mannerFull, pairwise~speech*gesture, adjust="
  ↳ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL
  ↳ asymp.UCL
dotted absent -0.46541506 0.04202377 NA -0.54778014
  ↳ -0.38304998
wiggly absent -0.22492188 0.04028818 NA -0.30388527
  ↳ -0.14595850
dotted present -0.12128568 0.04154759 NA -0.20271745
  ↳ -0.03985390
wiggly present -0.01267227 0.04020226 NA -0.09146724
  ↳ 0.06612271

Confidence level used: 0.95

$contrasts
contrast              estimate      SE df z.
  ↳ ratio p.value
dotted,absent - wiggly,absent -0.2404932 0.03935666 NA
  ↳ -6.111 <.0001
dotted,absent - dotted,present -0.3441294 0.04222057 NA
  ↳ -8.151 <.0001
dotted,absent - wiggly,present -0.4527428 0.04535124 NA
  ↳ -9.983 <.0001
wiggly,absent - dotted,present -0.1036362 0.04460096 NA
  ↳ -2.324 0.0927
wiggly,absent - wiggly,present -0.2122496 0.04224861 NA
  ↳ -5.024 <.0001
dotted,present - wiggly,present -0.1086134 0.03894116 NA
  ↳ -2.789 0.0271

P value adjustment: tukey method for comparing a family of
  ↳ 4 estimates

summary(wc.mannerFull)

```

```

Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: characterAdvantage ~ speech * gesture + (1 +
  ↳ speech + gesture | Participant) + (1 | MediaName)
Data: wc

AIC      BIC    logLik deviance df.resid
55245.4  55345.6 -27610.7  55221.4    31281

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.6174 -0.7349  0.0514  0.5535  3.3707

Random effects:
  Groups      Name              Variance Std.Dev. Corr
MediaName  (Intercept)          0.023847 0.15442
Participant (Intercept)          0.038569 0.19639
speechwiggly 0.008049 0.08972  -0.37
gesturepresent 0.016726 0.12933  -0.35  0.05
Residual                    0.334133 0.57804
Number of obs: 31293, groups:  MediaName, 158; Participant,
  ↳ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)          -0.46542      0.04202 -11.075
speechwiggly          0.24049      0.03936   6.111
gesturepresent        0.34413      0.04222   8.151
speechwiggly:gesturepresent -0.13188      0.05086  -2.593

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.511
gesturprsnt -0.514  0.404
spchwggly:g  0.306 -0.654 -0.602

#7100-7350
xc <- subset(df2, Timebin >= 7050 & Timebin <= 7350)

xc.null <- lmer(characterAdvantage ~ 1 + (1+gesture+speech|
  ↳ Participant) + (1|MediaName), data = xc, REML=FALSE)

xc.speech <- lmer(characterAdvantage ~ speech + (1+gesture+
  ↳ speech|Participant) + (1|MediaName), data = xc, REML=
  ↳ FALSE)

xc.gesture <- lmer(characterAdvantage ~ gesture + (1+
  ↳ gesture+speech|Participant) + (1|MediaName), data =
  ↳ xc, REML=FALSE)

xc.mannerFull <- lmer(characterAdvantage ~ speech*gesture +

```

```

  ↪ (1+gesture+speech|Participant) + (1|MediaName), data
  ↪ = xc, REML=FALSE)

xc.mannerSub <- lmer(characterAdvantage ~ speech + gesture
  ↪ + (1+gesture+speech|Participant) + (1|MediaName),
  ↪ data = xc, REML=FALSE)

anova(xc.null, xc.speech)
Data: xc
Models:
  xc.null: characterAdvantage ~ 1 + (1 + gesture + speech |
    ↪ Participant) + xc.null: (1 | MediaName)
  xc.speech: characterAdvantage ~ speech + (1 + gesture +
    ↪ speech | Participant) + xc.speech: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
xc.null    9 50727 50803 -25355    50709
xc.speech 10 50716 50799 -25348    50696 13.546    1
  ↪ 0.0002327 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

anova(xc.null, xc.gesture)
Data: xc
Models:
  xc.null: characterAdvantage ~ 1 + (1 + gesture + speech |
    ↪ Participant) + xc.null: (1 | MediaName)
  xc.gesture: characterAdvantage ~ gesture + (1 + gesture +
    ↪ speech | Participant) + xc.gesture: (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
xc.null    9 50727 50803 -25355    50709
xc.gesture 10 50689 50773 -25335    50669 40.067    1
  ↪ 2.454e-10 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

anova(xc.mannerSub, xc.mannerFull)
Data: xc
Models:
  xc.mannerSub: characterAdvantage ~ speech + gesture + (1
    ↪ + gesture + speech | xc.mannerSub: Participant)
    ↪ + (1 | MediaName)
  xc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
    ↪ gesture + speech | xc.mannerFull: Participant) +
    ↪ (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
xc.mannerSub 11 50667 50759 -25323    50645
xc.mannerFull 12 50666 50766 -25321    50642 3.4128    1
  ↪ 0.06469 .

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ' ' 1
↪ ' ' 1

lsmeans(xc.mannerFull, pairwise~speech*gesture, adjust="
  ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL
  ↪ asymp.UCL
dotted absent -0.396903920 0.03257874 NA -0.4607571
  ↪ -0.33305077
wiggly absent -0.193889902 0.03816545 NA -0.2686928
  ↪ -0.11908700
dotted present -0.117117792 0.03482284 NA -0.1853693
  ↪ -0.04886627
wiggly present -0.004562849 0.04000479 NA -0.0829708
  ↪ 0.07384510

Confidence level used: 0.95

$contrasts
contrast                estimate      SE df z
  ↪ .ratio p.value
dotted,absent - wiggly,absent -0.20301402 0.03912846 NA
  ↪ -5.188 <.0001
dotted,absent - dotted,present -0.27978613 0.03783450 NA
  ↪ -7.395 <.0001
dotted,absent - wiggly,present -0.39234107 0.04190584 NA
  ↪ -9.362 <.0001
wiggly,absent - dotted,present -0.07677211 0.04206632 NA
  ↪ -1.825 0.2615
wiggly,absent - wiggly,present -0.18932705 0.03785138 NA
  ↪ -5.002 <.0001
dotted,present - wiggly,present -0.11255494 0.03875800 NA
  ↪ -2.904 0.0193

P value adjustment: tukey method for comparing a family of
  ↪ 4 estimates

summary(xc.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: characterAdvantage ~ speech * gesture + (1 +
  ↪ gesture + speech | Participant) + (1 | MediaName
  ↪ )
Data: xc

AIC      BIC    logLik deviance df.resid
50665.8  50766.1 -25320.9  50641.8    31528

Scaled residuals:

```

Min	1Q	Median	3Q	Max
-2.9475	-0.7056	0.1251	0.5211	3.3024

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.021983	0.14827	
Participant	(Intercept)	0.015927	0.12620	
gesturepresent	0.008326	0.09125	-0.11	
speechwiggly	0.011267	0.10615	0.09	-0.01
Residual		0.285085	0.53393	

Number of obs: 31540, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t	value
(Intercept)		-0.39690	0.03258 -12.183
speechwiggly		0.20301	0.03913 5.188
gesturepresent		0.27979	0.03783 7.395
speechwiggly:gesturepresent		-0.09046	0.04870 -1.857

## Correlation of Fixed Effects:

(Intr)	spchwg	gstrpr
speechwggly	-0.446	
gesturprnt	-0.519	0.403
spchwgly:g	0.378	-0.630 -0.643

#7350-7650

yc &lt;- subset(df2, Timebin &gt;= 7350 &amp; Timebin &lt;= 7650)

yc.null <- lmer(characterAdvantage ~ 1 + (1+speech+gesture|  
 ↪ Participant) + (1|MediaName), data = yc, REML=FALSE)yc.speech <- lmer(characterAdvantage ~ speech + (1+speech+  
 ↪ gesture|Participant) + (1|MediaName), data = yc, REML  
 ↪ =FALSE)yc.gesture <- lmer(characterAdvantage ~ gesture + (1+speech  
 ↪ +gesture|Participant) + (1|MediaName), data = yc,  
 ↪ REML=FALSE)yc.mannerFull <- lmer(characterAdvantage ~ speech\*gesture +  
 ↪ (1+speech+gesture|Participant) + (1|MediaName), data  
 ↪ = yc, REML=FALSE)yc.mannerSub <- lmer(characterAdvantage ~ speech + gesture  
 ↪ + (1+speech+gesture|Participant) + (1|MediaName),  
 ↪ data = yc, REML=FALSE)

anova(yc.null, yc.speech)

Data: yc

Models:

yc.null: characterAdvantage ~ 1 + (1 + speech + gesture |  
 ↪ Participant) +

yc.null: (1 | MediaName)



```

yc.speech: characterAdvantage ~ speech + (1 + speech +
  ↪ gesture | Participant) +
  yc.speech:      (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
yc.null      9 45732 45807 -22857      45714
yc.speech 10 45730 45813 -22855      45710 4.6775      1
  ↪ 0.03056 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ '.' 1

anova(yc.null, yc.gesture)
Data: yc
Models:
  yc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) +  yc.null:      (1 | MediaName)
yc.gesture: characterAdvantage ~ gesture + (1 + speech +
  ↪ gesture | Participant) +  yc.gesture:      (1 |
  ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
yc.null      9 45732 45807 -22857      45714
yc.gesture 10 45707 45791 -22844      45687 27.001      1
  ↪ 2.033e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ '.' 1

anova(yc.mannerSub, yc.mannerFull)
Data: yc
Models:
  yc.mannerSub: characterAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | yc.mannerSub:      Participant)
    ↪ + (1 | MediaName)
yc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
  ↪ speech + gesture | yc.mannerFull:      Participant) +
  ↪ (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
yc.mannerSub  11 45693 45785 -22836      45671
yc.mannerFull 12 45690 45791 -22833      45666 4.4644      1
  ↪ 0.03461 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ '.' 1

lsmeans(yc.mannerFull, pairwise~speech*gesture, adjust="
  ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL
  ↪ asymp.UCL

```



```
dotted absent -0.25004372 0.02842845 NA -0.30576246
  ↪ -0.194324974
wiggly absent -0.07573298 0.03549594 NA -0.14530375
  ↪ -0.006162208
dotted present -0.01910860 0.03147351 NA -0.08079555
  ↪ 0.042578348
wiggly present 0.06664542 0.04085965 NA -0.01343802
  ↪ 0.146728867
```

Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z
↪ .ratio				p.value
dotted,absent - wiggly,absent	-0.17431074	0.03687593	NA	
↪	-4.727	<.0001		
dotted,absent - dotted,present	-0.23093512	0.03409276	NA	
↪	-6.774	<.0001		
dotted,absent - wiggly,present	-0.31668914	0.04328775	NA	
↪	-7.316	<.0001		
wiggly,absent - dotted,present	-0.05662438	0.03765933	NA	
↪	-1.504	0.4353		
wiggly,absent - wiggly,present	-0.14237840	0.03410315	NA	
↪	-4.175	0.0002		
dotted,present - wiggly,present	-0.08575403	0.03658900	NA	
↪	-2.344	0.0883		

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(yc.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: characterAdvantage ~ speech \* gesture + (1 +  
 ↪ speech + gesture | Participant) + (1 | MediaName  
 ↪ )

Data: yc

AIC	BIC	logLik	deviance	df.resid
45690.5	45790.6	-22833.2	45666.5	31180

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-3.4524	-0.3351	0.1214	0.4649	3.2108

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.01584	0.1259	
Participant	(Intercept)	0.01272	0.1128	
speechwiggly		0.01661	0.1289	-0.03
gesturepresent		0.01011	0.1005	-0.15 0.31

```

Residual                0.24746  0.4975
Number of obs: 31192, groups:  MediaName, 158; Participant,
  ↪ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)          -0.25004    0.02843  -8.796
speechwiggly          0.17431    0.03688   4.727
gesturerepresent      0.23094    0.03409   6.774
speechwiggly:gesturerepresent -0.08856    0.04162  -2.128

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.433
gesturprsnt -0.506  0.439
spchwggly:g  0.370 -0.571 -0.610

#7700-7950
zc <- subset(df2, Timebin >= 7650 & Timebin <= 7950)

zc.null <- lmer(characterAdvantage ~ 1 + (1+speech+gesture|
  ↪ Participant) + (1|MediaName), data = zc, REML=FALSE)
zc.speech <- lmer(characterAdvantage ~ speech + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = zc, REML
  ↪ =FALSE)
zc.gesture <- lmer(characterAdvantage ~ gesture + (1+speech
  ↪ +gesture|Participant) + (1|MediaName), data = zc,
  ↪ REML=FALSE)
zc.mannerFull <- lmer(characterAdvantage ~ speech*gesture +
  ↪ (1+speech+gesture|Participant) + (1|MediaName), data
  ↪ = zc, REML=FALSE)
zc.mannerSub <- lmer(characterAdvantage ~ speech + gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = zc, REML=FALSE)

anova(zc.null, zc.gesture)
Data: zc
Models:
  zc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
  ↪ Participant) +
  zc.null:      (1 | MediaName)
zc.gesture: characterAdvantage ~ gesture + (1 + speech +
  ↪ gesture | Participant) +
  zc.gesture:      (1 | MediaName)
Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
zc.null      9 44614 44689 -22298      44596
zc.gesture  10 44589 44672 -22285      44569 26.585      1
  ↪ 2.522e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

```

```

    ↪ ' ' 1

anova(zc.null, zc.speech)
Data: zc
Models:
  zc.null: characterAdvantage ~ 1 + (1 + speech + gesture |
    ↪ Participant) + zc.null:      (1 | MediaName)
  zc.speech: characterAdvantage ~ speech + (1 + speech +
    ↪ gesture | Participant) + zc.speech:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
zc.null    9 44614 44689 -22298    44596
zc.speech 10 44613 44696 -22296    44593 2.8818    1
    ↪ 0.08958 .

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

anova(zc.mannerSub, zc.mannerFull)
Data: zc
Models:
  zc.mannerSub: characterAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | zc.mannerSub: Participant) +
    ↪ (1 | MediaName)
  zc.mannerFull: characterAdvantage ~ speech * gesture + (1 +
    ↪ speech + gesture | zc.mannerFull:      Participant) +
    ↪ (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
zc.mannerSub  11 44581 44673 -22280    44559
zc.mannerFull 12 44582 44681 -22279    44558 1.6694    1
    ↪      0.1963

lsmeans(zc.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL
    ↪ asymp.UCL
dotted absent -0.08331696 0.02740685 NA -0.13703339
    ↪ -0.02960052
wiggly absent  0.04228378 0.03600260 NA -0.02828003
    ↪ 0.11284759
dotted present 0.10073597 0.03265324 NA  0.03673680
    ↪ 0.16473514
wiggly present 0.17911687 0.04217702 NA  0.09645142
    ↪ 0.26178232

Confidence level used: 0.95

$contrasts
contrast                                estimate      SE df z

```

```

↪ .ratio p.value
dotted,absent - wiggly,absent -0.12560074 0.03584830 NA
↪ -3.504 0.0026
dotted,absent - dotted,present -0.18405292 0.03090626 NA
↪ -5.955 <.0001
dotted,absent - wiggly,present -0.26243383 0.04167290 NA
↪ -6.297 <.0001
wiggly,absent - dotted,present -0.05845219 0.03741234 NA
↪ -1.562 0.4003
wiggly,absent - wiggly,present -0.13683309 0.03090598 NA
↪ -4.427 0.0001
dotted,present - wiggly,present -0.07838090 0.03562682 NA
↪ -2.200 0.1232

```

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(zc.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: characterAdvantage ~ speech \* gesture + (1 +  
 ↪ speech + gesture | Participant) + (1 | MediaName  
 ↪ )

Data: zc

AIC	BIC	logLik	<b>deviance</b>	<b>df.resid</b>
44581.5	44681.4	-22278.8	44557.5	30313

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-3.8258	-0.3401	0.0271	0.3898	2.5519

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.011811	0.10868	
Participant	(Intercept)	0.014306	0.11961	
speechwiggly	0.021106	0.14528	-0.06	
gesturepresent	0.009912	0.09956	0.06	0.21
Residual		0.248910	0.49891	

Number of obs: 30325, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	<b>t</b>	value
(Intercept)		-0.08332	0.02741 -3.040
speechwiggly		0.12560	0.03585 3.504
gesturepresent		0.18405	0.03091 5.955
speechwiggly:gesturepresent		-0.04722	0.03645 -1.296

Correlation of Fixed Effects:

(Intr) spchwg gstrpr

```

speechwggly -0.377
gesturprsnt -0.378 0.379
spchwggly:g 0.336 -0.514 -0.590

#D/O vs Distractor

#3 create an array that includes only looks to D1 and D2
ad <- subset(e, AOI == "AOI.D1.Hit" | AOI == "AOI.D2.Hit")

#4. create a column containing the average proportion of
  ↳ looks to relevabt AOI per item per participant

bd <- ddply(ad,.(MediaName, Participant, Timebin),transform
  ↳ ,DO_DisProp=mean(prop))

#5.create a new df containing only those columns of
  ↳ interest including the new DO_DisProp column

cd <- bd[, c("Participant", "MediaName", "condition", "
  ↳ Timebin", "gesture", "speech", "DO_DisProp")]

cd$DO_DisProp <- ifelse(cd$DO_DisProp >0, 1, 0)

#Create a df that includes only looks to D3
dd<- e[e$AOI == "AOI.D3.Hit",]

#4. create a column containing the average proportion of
  ↳ looks to relevabt AOI per item per participant

dd <- ddply(dd,.(MediaName, Participant, Timebin),transform
  ↳ ,DisProp=mean(prop))

#5.create a new df containing only those columns of
  ↳ interest including the new DisProp column

ed <- dd[, c("Participant", "MediaName", "condition", "
  ↳ Timebin", "gesture", "speech", "DisProp")]

#Merge dfs into distractor df

df3 <- merge(cd, ed, all=T)

#Subtract the scores associated with d3 from d1 and D2

df3$competitorAdvantage <- df3$DO_DisProp - df3$DisProp

write.csv(df3, "competitorData.csv")
df3 <- read.csv("competitorData.csv", header = T)
scatterDis <- ggplot(df3, aes(Timebin, competitorAdvantage,

```

```

  ↪ colour = condition))

scatterDis + stat_summary(fun.y = mean, geom = "point") +
  ↪ stat_summary(fun.y = mean, geom = "line", aes (group
  ↪ = condition)) + labs(x = "Time(ms)", y = "Competitor_
  ↪ Advantage", colour = "condition") + ggtitle("
  ↪ Competitor_Advantage") + annotate("rect", xmin =
  ↪ 5270, xmax = 5540, ymin = 0.0, ymax = 0.5, alpha =
  ↪ .2)+ annotate("rect", xmin = 6950, xmax = 7500, ymin
  ↪ = 0.0, ymax = 0.5, alpha = .2) + scale_color_manual("
  ↪ Condition", breaks=c("condition1","condition2", "
  ↪ condition3", "condition4"), labels=c("+Speech;_+
  ↪ Gesture", "+Speech;_-Gesture", "-Speech;_+Gesture", "
  ↪ -Speech;_-Gesture"), values = c("red", "blue", "green
  ↪ ", "darkgreen")) + theme(legend.text=element_text(
  ↪ size=12), axis.text=element_text(size=12), axis.title
  ↪ =element_text(size=14,face="bold"), legend.title=
  ↪ element_text(size=14), plot.title=element_text(size
  ↪ =16)) + annotate("text", x = 5100, y = 0.3, label = "
  ↪ mannerDOP") + annotate("text", x = 8000, y = 0.40,
  ↪ label = "groundDOP")

```

```
#5250-5550
```

```

rd <- subset(df3, Timebin >= 5250 & Timebin <= 5550)

rd.null <- lmer(competitorAdvantage ~ 1 + (1+speech+gesture
  ↪ |Participant) + (1|MediaName), data = rd, REML=FALSE)

rd.speech <- lmer(competitorAdvantage ~ speech + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = rd, REML
  ↪ =FALSE)

rd.gesture <- lmer(competitorAdvantage ~ gesture + (1+
  ↪ speech+gesture|Participant) + (1|MediaName), data =
  ↪ rd, REML=FALSE)

rd.mannerFull <- lmer(competitorAdvantage ~ speech*gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = rd, REML=FALSE)

rd.mannerSub <- lmer(competitorAdvantage ~ speech + gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = rd, REML=FALSE)

```

```
anova(rd.null, rd.gesture)
```

```
Data: rd
```

```
Models:
```

```

  rd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
  ↪ | Participant) + rd.null:      (1 | MediaName)
rd.gesture: competitorAdvantage ~ gesture + (1 + speech +

```

```

    ↪ gesture | Participant) + rd.gesture: (1 | MediaName
    ↪ )
Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
rd.null      9 8970.8 9052.1 -4476.4   8952.8
rd.gesture 10 8972.4 9062.8 -4476.2   8952.4 0.3543      1
    ↪      0.5517

anova(rd.null, rd.speech)
Data: rd
Models:
  rd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
    ↪ | Participant) + rd.null:      (1 | MediaName)
  rd.speech: competitorAdvantage ~ speech + (1 + speech +
    ↪ gesture | Participant) + rd.speech:      (1 |
    ↪ MediaName)
Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
rd.null      9 8970.8 9052.1 -4476.4   8952.8
rd.speech 10 8972.7 9063.1 -4476.4   8952.7 0.0385      1
    ↪      0.8445

anova(rd.mannerSub, rd.mannerFull)
Data: rd
Models:
  rd.mannerSub: competitorAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | rd.mannerSub:      Participant
    ↪ ) + (1 | MediaName)
  rd.mannerFull: competitorAdvantage ~ speech * gesture + (1
    ↪ + speech + gesture | rd.mannerFull:      Participant)
    ↪ + (1 | MediaName)
Df      AIC      BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
rd.mannerSub 11 8974.4 9073.8 -4476.2   8952.4
rd.mannerFull 12 8974.9 9083.4 -4475.5   8950.9 1.4773
    ↪      1      0.2242

summary(rd.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: competitorAdvantage ~ speech * gesture + (1 +
    ↪ speech + gesture |      Participant) + (1 | MediaName
    ↪ )
Data: rd

AIC      BIC  logLik deviance df.resid
8974.9    9083.4 -4475.5   8950.9    62124

Scaled residuals:
  Min      1Q  Median      3Q      Max
-4.8871 -0.2454 -0.0950  0.0487  4.5587

Random effects:
  Groups      Name      Variance Std.Dev.  Corr

```



```

MediaName (Intercept) 0.002075 0.04555
Participant (Intercept) 0.002279 0.04774
speechwiggly 0.002081 0.04562 -0.29
gesturepresent 0.001266 0.03558 -0.61 -0.10
Residual 0.066880 0.25861
Number of obs: 62136, groups: MediaName, 158; Participant,
  ↪ 35
    
```

Fixed **effects**:

	Estimate	Std. Error	t	value
(Intercept)	0.035152	0.011130	3.158	
speechwiggly	0.007451	0.013267	0.562	
gesturepresent	0.003455	0.012292	0.281	
speechwiggly:gesturepresent	-0.018376	0.015093	-1.218	

Correlation of Fixed Effects:

```

(Intr) spchwg gstrpr
speechwggly -0.518
gesturprsnt -0.642 0.324
spchwggly:g 0.343 -0.575 -0.613
    
```

```

lsmeans(rd.mannerFull, pairwise~speech*gesture, adjust="
  ↪ tukey")
    
```

```

$lsmeans
speech gesture lsmean SE df asymp.LCL asymp
  ↪ .UCL
dotted absent 0.03515178 0.011129939 NA 0.013337502
  ↪ 0.05696606
wiggly absent 0.04260309 0.012125806 NA 0.018836947
  ↪ 0.06636923
dotted present 0.03860652 0.009964054 NA 0.019077330
  ↪ 0.05813571
wiggly present 0.02768185 0.010637107 NA 0.006833505
  ↪ 0.04853020
    
```

Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df
↪ z.ratio p.value			
dotted,absent - wiggly,absent	-0.007451308	0.01326729	NA
↪ -0.562 0.9434			
dotted,absent - dotted,present	-0.003454736	0.01229219	NA
↪ -0.281 0.9923			
dotted,absent - wiggly,present	0.007469930	0.01423573	NA
↪ 0.525 0.9531			
wiggly,absent - dotted,present	0.003996572	0.01488256	NA
↪ 0.269 0.9932			
wiggly,absent - wiggly,present	0.014921238	0.01230302	NA
↪ 1.213 0.6187			



```
dotted,present - wiggly,present  0.010924666 0.01317047 NA
  ↪ 0.829 0.8405

P value adjustment: tukey method for comparing a family of
  ↪ 4 estimates

#5550 - 5850
sd<- subset(df3, Timebin >= 5550 & Timebin <= 5850)

sd.null <- lmer(competitorAdvantage ~ 1 + (1+gesture+speech
  ↪ |Participant) + (1|MediaName), data = sd, REML=FALSE)

sd.speech <- lmer(competitorAdvantage ~ speech + (1+gesture
  ↪ +speech|Participant) + (1|MediaName), data = sd, REML
  ↪ =FALSE)

sd.gesture <- lmer(competitorAdvantage ~ gesture + (1+
  ↪ gesture+speech|Participant) + (1|MediaName), data =
  ↪ sd, REML=FALSE)

sd.mannerFull <- lmer(competitorAdvantage ~ speech*gesture
  ↪ + (1+gesture+speech|Participant) + (1|MediaName),
  ↪ data = sd, REML=FALSE)

sd.mannerSub <- lmer(competitorAdvantage ~ speech + gesture
  ↪ + (1+gesture+speech|Participant) + (1|MediaName),
  ↪ data = sd, REML=FALSE)

anova(sd.null, sd.gesture)
Data: sd
Models:
  sd.null: competitorAdvantage ~ 1 + (1 + gesture + speech
  ↪ | Participant) + sd.null:      (1 | MediaName)
sd.gesture: competitorAdvantage ~ gesture + (1 + gesture +
  ↪ speech | Participant) + sd.gesture:      (1 |
  ↪ MediaName)
Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
sd.null      9 48082 48163 -24032      48064
sd.gesture 10 48083 48173 -24031      48063 1.0739      1
  ↪      0.3001

anova(sd.null, sd.speech)
Data: sd
Models:
  sd.null: competitorAdvantage ~ 1 + (1 + gesture + speech
  ↪ | Participant) + sd.null:      (1 | MediaName)
sd.speech: competitorAdvantage ~ speech + (1 + gesture +
  ↪ speech | Participant) + sd.speech:      (1 | MediaName
  ↪ )
Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
```

```

sd.null      9 48082 48163 -24032      48064
sd.speech 10 48084 48174 -24032      48064 0.1843      1
  ↪ 0.6677

anova(sd.mannerSub, sd.mannerFull)
Data: sd
Models:
  sd.mannerSub: competitorAdvantage ~ speech + gesture + (1
  ↪ + gesture + speech | sd.mannerSub:      Participant
  ↪ ) + (1 | MediaName)
  sd.mannerFull: competitorAdvantage ~ speech * gesture + (1
  ↪ + gesture + speech | sd.mannerFull:      Participant)
  ↪ + (1 | MediaName)
Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
sd.mannerSub  11 48084 48184 -24031      48062
sd.mannerFull 12 48086 48195 -24031      48062 0.0079      1
  ↪      0.9293

lsmeans(sd.mannerFull, pairwise~speech*gesture, adjust="
  ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL  asymp.
  ↪ UCL
dotted absent  0.08894445 0.01771891 NA 0.05421602
  ↪ 0.12367287
wiggly absent  0.08230335 0.01764628 NA 0.04771728
  ↪ 0.11688943
dotted present 0.07048345 0.01564915 NA 0.03981168
  ↪ 0.10115522
wiggly present 0.06594664 0.01470294 NA 0.03712940
  ↪ 0.09476387

Confidence level used: 0.95

$contrasts
contrast                                     estimate      SE df z
  ↪ .ratio p.value
dotted,absent - wiggly,absent  0.006641094 0.01903410 NA
  ↪ 0.349 0.9854
dotted,absent - dotted,present 0.018460995 0.02082255 NA
  ↪ 0.887 0.8118
dotted,absent - wiggly,present 0.022997810 0.02203166 NA
  ↪ 1.044 0.7236
wiggly,absent - dotted,present 0.011819901 0.02319445 NA
  ↪ 0.510 0.9568
wiggly,absent - wiggly,present 0.016356716 0.02082675 NA
  ↪ 0.785 0.8611
dotted,present - wiggly,present 0.004536815 0.01885393 NA
  ↪ 0.241 0.9951

```

```

P value adjustment: tukey method for comparing a family of
  ↪ 4 estimates

summary(sd.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: competitorAdvantage ~ speech * gesture + (1 +
  ↪ gesture + speech | Participant) + (1 | MediaName)
Data: sd

AIC      BIC    logLik deviance df.resid
48086.5  48195.1 -24031.2  48062.5    62902

Scaled residuals:
  Min      1Q  Median      3Q      Max
-4.147 -0.330 -0.126  0.050  3.093

Random effects:
  Groups      Name                Variance Std.Dev. Corr
MediaName (Intercept)    0.005230 0.07232
Participant (Intercept) 0.005931 0.07701
gesturepresent 0.005248 0.07244 -0.68
speechwiggly 0.002649 0.05147 -0.35 -0.12
Residual                0.124208 0.35243
Number of obs: 62914, groups:  MediaName, 158; Participant,
  ↪ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)          0.088944  0.017719  5.020
speechwiggly        -0.006641  0.019034 -0.349
gesturepresent      -0.018461  0.020823 -0.887
speechwiggly:gesturepresent 0.002104  0.023701  0.089

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.541
gesturprsnt -0.681  0.325
spchwggly:g  0.339 -0.630 -0.569

#5900-6150
td <- subset(df3, Timebin >= 5850 & Timebin <= 6150)

td.null <- lmer(competitorAdvantage ~ 1 + (1+gesture+speech
  ↪ |Participant) + (1|MediaName), data = td, REML=FALSE)

td.speech <- lmer(competitorAdvantage ~ speech + (1+gesture
  ↪ +speech|Participant) + (1|MediaName), data = td, REML
  ↪ =FALSE)

td.gesture <- lmer(competitorAdvantage ~ gesture + (1+

```

```

↪ gesture+speech|Participant) + (1|MediaName), data =
↪ td, REML=FALSE)

td.mannerFull <- lmer(competitorAdvantage ~ speech*gesture
↪ + (1+gesture+speech|Participant) + (1|MediaName),
↪ data = td, REML=FALSE)

td.mannerSub <- lmer(competitorAdvantage ~ speech + gesture
↪ + (1+gesture+speech|Participant) + (1|MediaName),
↪ data = td, REML=FALSE)

anova(td.null, td.gesture)
Data: td
Models:
  td.null: competitorAdvantage ~ 1 + (1 + gesture + speech
↪ | Participant) + td.null:      (1 | MediaName)
  td.gesture: competitorAdvantage ~ gesture + (1 + gesture +
↪ speech | Participant) + td.gesture:      (1 |
↪ MediaName)
Df  AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
td.null    9 72063 72145 -36023    72045
td.gesture 10 72058 72148 -36019    72038 7.2383    1
↪ 0.007136 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
↪ ' ' 1

anova(td.null, td.speech)
Data: td
Models:
  td.null: competitorAdvantage ~ 1 + (1 + gesture + speech
↪ | Participant) + td.null:      (1 | MediaName)
  td.speech: competitorAdvantage ~ speech + (1 + gesture +
↪ speech | Participant) + td.speech: (1 | MediaName)
Df  AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
td.null    9 72063 72145 -36023    72045
td.speech 10 72065 72155 -36023    72045 0.1081    1
↪ 0.7423

anova(td.mannerSub, td.mannerFull)
Data: td
Models:
  td.mannerSub: competitorAdvantage ~ speech + gesture + (1
↪ + gesture + speech | td.mannerSub:      Participant
↪ ) + (1 | MediaName)
  td.mannerFull: competitorAdvantage ~ speech * gesture + (1
↪ + gesture + speech | td.mannerFull:      Participant)
↪ + (1 | MediaName)
Df  AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
td.mannerSub 11 72060 72159 -36019    72038

```

```

td.mannerFull 12 72059 72167 -36017      72035 2.8831      1
  ↳ 0.08952 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↳ '.' 1

lsmeans(td.mannerFull, pairwise~speech*gesture, adjust="
  ↳ tukey")
$lsmeans
speech gesture      lsmean          SE df  asymp.LCL asymp.
  ↳ UCL
dotted absent  0.17055201 0.02201269 NA 0.12740793
  ↳ 0.2136961
wiggly absent  0.13162154 0.02151614 NA 0.08945068
  ↳ 0.1737924
dotted present 0.08434227 0.01962743 NA 0.04587322
  ↳ 0.1228113
wiggly present 0.09985142 0.02039375 NA 0.05988040
  ↳ 0.1398224

Confidence level used: 0.95

$contrasts
contrast                                estimate          SE df z
  ↳ .ratio p.value
dotted,absent - wiggly,absent    0.03893047 0.02693064 NA
  ↳ 1.446 0.4708
dotted,absent - dotted,present  0.08620974 0.02653215 NA
  ↳ 3.249 0.0064
dotted,absent - wiggly,present  0.07070059 0.03106843 NA
  ↳ 2.276 0.1037
wiggly,absent - dotted,present  0.04727927 0.02934127 NA
  ↳ 1.611 0.3721
wiggly,absent - wiggly,present  0.03177012 0.02652485 NA
  ↳ 1.198 0.6283
dotted,present - wiggly,present -0.01550915 0.02668231 NA
  ↳ -0.581 0.9378

P value adjustment: tukey method for comparing a family of
  ↳ 4 estimates

summary(td.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: competitorAdvantage ~ speech * gesture + (1 +
  ↳ gesture + speech | Participant) + (1 | MediaName
  ↳ )
Data: td

AIC      BIC    logLik deviance df.resid
72058.8  72167.3 -36017.4  72034.8    62454

```

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-3.6659	-0.3756	-0.1657	0.0562	3.2921

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.009591	0.09793	
Participant	(Intercept)	0.007803	0.08833	
gesturepresent		0.006666	0.08164	-0.69
speechwiggly		0.007183	0.08475	-0.53 0.13
Residual		0.183185	0.42800	

Number of obs: 62466, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t	value
(Intercept)		0.17055	0.02201 7.748
speechwiggly		-0.03893	0.02693 -1.446
gesturepresent		-0.08621	0.02653 -3.249
speechwiggly:gesturepresent		0.05444	0.03192 1.706

## Correlation of Fixed Effects:

(Intr)	spchwg	gstrpr
speechwggly	-0.630	
gesturprsnt	-0.688	0.398
spchwggly:g	0.367	-0.600 -0.602

#6150-6450

ud <- **subset**(df3, Timebin >= 6150 & Timebin <= 6450)ud.**null** <- lmer(competitorAdvantage ~ 1 + (1+speech+gesture  
 ↪ |Participant) + (1|MediaName), **data** = ud, REML=FALSE)ud.speech <- lmer(competitorAdvantage ~ speech + (1+speech+  
 ↪ gesture|Participant) + (1|MediaName), **data** = ud, REML  
 ↪ =FALSE)ud.gesture <- lmer(competitorAdvantage ~ gesture + (1+  
 ↪ speech+gesture|Participant) + (1|MediaName), **data** =  
 ↪ ud, REML=FALSE)ud.mannerFull <- lmer(competitorAdvantage ~ speech\*gesture  
 ↪ + (1+speech+gesture|Participant) + (1|MediaName),  
 ↪ **data** = ud, REML=FALSE)ud.mannerSub <- lmer(competitorAdvantage ~ speech + gesture  
 ↪ + (1+speech+gesture|Participant) + (1|MediaName),  
 ↪ **data** = ud, REML=FALSE)

```

anova(ud.null, ud.speech)
Data: ud
Models:
  ud.null: competitorAdvantage ~ 1 + (1 + speech + gesture
    ↪ | Participant) + ud.null:      (1 | MediaName)
  ud.speech: competitorAdvantage ~ speech + (1 + speech +
    ↪ gesture | Participant) + ud.speech:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
ud.null      9 79195 79276 -39589      79177
ud.speech    10 79190 79280 -39585      79170 7.4516      1
    ↪ 0.006338 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

anova(ud.null, ud.gesture)
Data: ud
Models:
  ud.null: competitorAdvantage ~ 1 + (1 + speech + gesture
    ↪ | Participant) + ud.null:      (1 | MediaName)
  ud.gesture: competitorAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + ud.gesture:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
ud.null      9 79195 79276 -39589      79177
ud.gesture   10 79178 79269 -39579      79158 18.729      1
    ↪ 1.507e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

anova(ud.mannerSub, ud.mannerFull)
Data: ud
Models:
  ud.mannerSub: competitorAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | ud.mannerSub:      Participant
    ↪ ) + (1 | MediaName)
  ud.mannerFull: competitorAdvantage ~ speech * gesture + (1
    ↪ + speech + gesture | ud.mannerFull:      Participant)
    ↪ + (1 | MediaName)
Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
ud.mannerSub  11 79170 79270 -39574      79148
ud.mannerFull 12 79167 79275 -39571      79143 5.2502      1
    ↪ 0.02194 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

lsmeans(ud.mannerFull, pairwise~speech*gesture, adjust="

```

```

    ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL asymp.
    ↪ UCL
dotted absent  0.29283044 0.02885716 NA 0.23627144
    ↪ 0.3493894
wiggly absent  0.16035312 0.02649203 NA 0.10842970
    ↪ 0.2122765
dotted present 0.11219041 0.02477037 NA 0.06364138
    ↪ 0.1607394
wiggly present 0.07968286 0.02358982 NA 0.03344767
    ↪ 0.1259181

Confidence level used: 0.95

$contrasts
contrast                                     estimate      SE df z.
    ↪ ratio p.value
dotted,absent - wiggly,absent  0.13247731 0.03320331 NA
    ↪ 3.990 0.0004
dotted,absent - dotted,present 0.18064002 0.03445187 NA
    ↪ 5.243 <.0001
dotted,absent - wiggly,present 0.21314757 0.03762263 NA
    ↪ 5.665 <.0001
wiggly,absent - dotted,present 0.04816271 0.03560435 NA
    ↪ 1.353 0.5292
wiggly,absent - wiggly,present 0.08067026 0.03445442 NA
    ↪ 2.341 0.0888
dotted,present - wiggly,present 0.03250755 0.03284039 NA
    ↪ 0.990 0.7552

P value adjustment: tukey method for comparing a family of
    ↪ 4 estimates

summary(ud.mannerFull)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: competitorAdvantage ~ speech * gesture + (1 +
    ↪ speech + gesture | Participant) + (1 | MediaName
    ↪ )
Data: ud

AIC      BIC  logLik deviance df.resid
79166.9  79275.3 -39571.4  79142.9    62018

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.8387 -0.4113 -0.1675  0.1292  3.5216

Random effects:
  Groups      Name      Variance Std.Dev. Corr

```



```

MediaName (Intercept) 0.017952 0.13399
Participant (Intercept) 0.012435 0.11151
speechwiggly 0.005278 0.07265 -0.61
gesturepresent 0.008639 0.09294 -0.77 0.19
Residual 0.206881 0.45484
Number of obs: 62030, groups: MediaName, 158; Participant,
  ↪ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept) 0.29283 0.02886 10.148
speechwiggly -0.13248 0.03320 -3.990
gesturepresent -0.18064 0.03445 -5.243
speechwiggly:gesturepresent 0.09997 0.04326 2.311

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.644
gesturprnt -0.707 0.447
spchwggly:g 0.380 -0.660 -0.628

#6450-6750
vd <- subset(df3, Timebin >= 6450 & Timebin <= 6750)

vd.null <- lmer(competitorAdvantage ~ 1 + (1+speech+gesture
  ↪ |Participant) + (1|MediaName), data = vd, REML=FALSE)

vd.speech <- lmer(competitorAdvantage ~ speech + (1+speech+
  ↪ gesture|Participant) + (1|MediaName), data = vd, REML
  ↪ =FALSE)

vd.gesture <- lmer(competitorAdvantage ~ gesture + (1+
  ↪ speech+gesture|Participant) + (1|MediaName), data =
  ↪ vd, REML=FALSE)

vd.mannerFull <- lmer(competitorAdvantage ~ speech*gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = vd, REML=FALSE)

vd.mannerSub <- lmer(competitorAdvantage ~ speech + gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = vd, REML=FALSE)

anova(vd.null, vd.speech)
Data: vd
Models:
  vd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
  ↪ | Participant) +
  vd.null: (1 | MediaName)
vd.speech: competitorAdvantage ~ speech + (1 + speech +

```

```

    ↪ gesture | Participant) +
    vd.speech:      (1 | MediaName)
Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
vd.null      9 86001 86083 -42992      85983
vd.speech 10 85993 86083 -42986      85973 10.594      1
    ↪ 0.001134 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ '.' 1

anova(vd.null, vd.gesture)
Data: vd
Models:
  vd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
    ↪ | Participant) + vd.null:      (1 | MediaName)
  vd.gesture: competitorAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + vd.gesture:      (1 |
    ↪ MediaName)
Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
vd.null      9 86001 86083 -42992      85983
vd.gesture 10 85972 86062 -42976      85952 31.332      1
    ↪ 2.175e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ '.' 1

anova(vd.mannerSub, vd.mannerFull)
Data: vd
Models:
  vd.mannerSub: competitorAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | vd.mannerSub:
    ↪ Participant) + (1 | MediaName)
  vd.mannerFull: competitorAdvantage ~ speech * gesture + (1
    ↪ + speech + gesture | vd.mannerFull:      Participant)
    ↪ + (1 | MediaName)
Df  AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
vd.mannerSub 11 85957 86056 -42967      85935
vd.mannerFull 12 85948 86057 -42962      85924 10.404      1
    ↪ 0.001257 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ '.' 1

lsmeans(vd.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df asymp.LCL asymp.
    ↪ UCL
dotted absent 0.38263932 0.02924814 NA 0.32531401
    ↪ 0.4399646

```

```
wiggly absent 0.19088286 0.02886363 NA 0.13431118
  ↳ 0.2474545
dotted present 0.11212656 0.02468151 NA 0.06375169
  ↳ 0.1605014
wiggly present 0.07007828 0.02557064 NA 0.01996074
  ↳ 0.1201958
```

Confidence level used: 0.95

**\$contrasts**

contrast	estimate	SE	df	z.
↳ ratio p.value				
dotted,absent - wiggly,absent	0.19175646	0.03529679	NA	
↳ 5.433 <.0001				
dotted,absent - dotted,present	0.27051275	0.03687891	NA	
↳ 7.335 <.0001				
dotted,absent - wiggly,present	0.31256104	0.04022812	NA	
↳ 7.770 <.0001				
wiggly,absent - dotted,present	0.07875630	0.03851987	NA	
↳ 2.045 0.1717				
wiggly,absent - wiggly,present	0.12080458	0.03688434	NA	
↳ 3.275 0.0058				
dotted,present - wiggly,present	0.04204829	0.03491904	NA	
↳ 1.204 0.6242				

P value adjustment: tukey method **for** comparing a **family** of  
↳ 4 estimates

**summary(vd.mannerFull)**

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
Formula: competitorAdvantage ~ speech \* gesture + (1 +  
↳ speech + gesture | Participant) + (1 | MediaName)  
Data: vd

AIC	BIC	logLik	deviance	df.resid
85948.5	86056.9	-42962.2	85924.5	61722

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-3.9021	-0.4080	-0.1572	0.4068	3.0056

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.019975	0.14133	
Participant	(Intercept)	0.011339	0.10649	
speechwiggly	0.006519	0.08074	-0.43	
gesturepresent	0.010969	0.10474	-0.86	0.14
Residual		0.232326	0.48200	

Number of obs: 61734, groups: MediaName, 158; Participant,  
↳ 35

Fixed **effects**:

	Estimate	Std. Error	t value
(Intercept)	0.38264	0.02925	13.083
speechwiggly	-0.19176	0.03530	-5.433
gesturepresent	-0.27051	0.03688	-7.335
speechwiggly:gesturepresent	0.14971	0.04565	3.280

Correlation of Fixed Effects:

	(Intr)	spchwg	gstrpr
speechwggly	-0.614		
gesturprsnt	-0.745	0.431	
spchwggly:g	0.395	-0.655	-0.619

#6800-7050

wd <- **subset**(df3, Timebin >= 6750 & Timebin <= 7050)

wd.**null** <- lmer(competitorAdvantage ~ 1 + (1+speech+gesture  
 ↪ |Participant) + (1|MediaName), **data** = wd, REML=FALSE)

wd.speech <- lmer(competitorAdvantage ~ speech + (1+speech+  
 ↪ gesture|Participant) + (1|MediaName), **data** = wd, REML  
 ↪ =FALSE)

wd.gesture <- lmer(competitorAdvantage ~ gesture + (1+  
 ↪ speech+gesture|Participant) + (1|MediaName), **data** =  
 ↪ wd, REML=FALSE)

wd.mannerFull <- lmer(competitorAdvantage ~ speech\*gesture  
 ↪ + (1+speech+gesture|Participant) + (1|MediaName),  
 ↪ **data** = wd, REML=FALSE)

wd.mannerSub <- lmer(competitorAdvantage ~ speech + gesture  
 ↪ + (1+speech+gesture|Participant) + (1|MediaName),  
 ↪ **data** = wd, REML=FALSE)

**anova**(wd.**null**, wd.speech)

Data: wd

Models:

wd.**null**: competitorAdvantage ~ 1 + (1 + speech + gesture  
 ↪ | Participant) + wd.**null**: (1 | MediaName)

wd.speech: competitorAdvantage ~ speech + (1 + speech +  
 ↪ gesture | Participant) + wd.speech: (1 |  
 ↪ MediaName)

Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
wd. <b>null</b>	9	90046	90128	-45014	90028		
wd.speech	10	90041	90131	-45010	90021	7.39	1
							↪ 0.006559 **

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1  
 ↪ ‘ ’ 1

**anova**(wd.**null**, wd.gesture)

```

Data: wd
Models:
  wd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
    ↪ | Participant) + wd.null:      (1 | MediaName)
wd.gesture: competitorAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + wd.gesture:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
wd.null      9 90046 90128 -45014      90028
wd.gesture  10 90022 90113 -45001      90002 25.975      1
    ↪ 3.459e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

anova(wd.mannerSub, wd.mannerFull)
Data: wd
Models:
  wd.mannerSub: competitorAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | wd.mannerSub:      Participant
    ↪ ) + (1 | MediaName)
wd.mannerFull: competitorAdvantage ~ speech * gesture + (1
    ↪ + speech + gesture | wd.mannerFull:      Participant)
    ↪ + (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
wd.mannerSub  11 90013 90112 -44995      89991
wd.mannerFull 12 90006 90115 -44991      89982 8.4795      1
    ↪ 0.003592 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
    ↪ ' ' 1

lsmeans(wd.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean          SE df  asymp.LCL  asymp.UCL
dotted absent  0.3661935 0.02666933 NA 0.31392253 0.4184644
wiggly absent  0.2066483 0.02673922 NA 0.15424038 0.2590562
dotted present 0.1306450 0.02506014 NA 0.08152799 0.1797620
wiggly present 0.0989395 0.02616233 NA 0.04766227 0.1502167

Confidence level used: 0.95

$contrasts
contrast
    ↪ ratio p.value
dotted,absent - wiggly,absent  0.15954517 0.03470321 NA
    ↪ 4.597 <.0001
dotted,absent - dotted,present 0.23554848 0.03594464 NA
    ↪ 6.553 <.0001

```

```
dotted,absent - wiggly,present  0.26725395  0.03999930  NA
  ↪ 6.681  <.0001
wiggly,absent - dotted,present  0.07600331  0.03865204  NA
  ↪ 1.966  0.2008
wiggly,absent - wiggly,present  0.10770878  0.03595714  NA
  ↪ 2.995  0.0146
dotted,present - wiggly,present  0.03170547  0.03435877  NA
  ↪ 0.923  0.7927
```

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(wd.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: competitorAdvantage ~ speech \* gesture + (1 +  
 ↪ speech + gesture | Participant) + (1 | MediaName  
 ↪ )

Data: wd

```
AIC      BIC    logLik deviance df.resid
90006.3  90114.8 -44991.1  89982.3    62574
```

Scaled **residuals**:

```
  Min      1Q  Median      3Q      Max
-3.6744 -0.4417 -0.1654  0.6259  2.4894
```

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.017904	0.13381	
Participant	(Intercept)	0.008152	0.09029	
speechwiggly	0.008732	0.09344	-0.51	
gesturepresent	0.012218	0.11054	-0.74	0.09
Residual		0.243321	0.49328	

Number of obs: 62586, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t value
(Intercept)	0.36619	0.02667 13.731
speechwiggly	-0.15955	0.03470 -4.597
gesturepresent	-0.23555	0.03594 -6.553
speechwiggly:gesturepresent	0.12784	0.04331 2.952

Correlation of Fixed Effects:

```
(Intr) spchwg gstrpr
speechwggly -0.649
gesturprsnt -0.717  0.402
spchwggly:g  0.411 -0.632 -0.602
```

#7100-7350

```

xd <- subset(df3, Timebin >= 7050 & Timebin <= 7350)

xd.null <- lmer(competitorAdvantage ~ 1 + (1+gesture+speech
  ↪ |Participant) + (1|MediaName), data = xd, REML=FALSE)

xd.speech <- lmer(competitorAdvantage ~ speech + (1+gesture
  ↪ +speech|Participant) + (1|MediaName), data = xd, REML
  ↪ =FALSE)

xd.gesture <- lmer(competitorAdvantage ~ gesture + (1+
  ↪ gesture+speech|Participant) + (1|MediaName), data =
  ↪ xd, REML=FALSE)

xd.mannerFull <- lmer(competitorAdvantage ~ speech*gesture
  ↪ + (1+gesture+speech|Participant) + (1|MediaName),
  ↪ data = xd, REML=FALSE)

xd.mannerSub <- lmer(competitorAdvantage ~ speech + gesture
  ↪ + (1+gesture+speech|Participant) + (1|MediaName),
  ↪ data = xd, REML=FALSE)

anova(xd.null, xd.speech)
Data: xd
Models:
  xd.null: competitorAdvantage ~ 1 + (1 + gesture + speech
    ↪ | Participant) + xd.null:      (1 | MediaName)
  xd.speech: competitorAdvantage ~ speech + (1 + gesture +
    ↪ speech | Participant) + xd.speech:      (1 | MediaName
    ↪ )
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
xd.null      9 86560 86641 -43271      86542
xd.speech  10 86560 86651 -43270      86540 1.5661      1
  ↪ 0.2108

anova(xd.null, xd.gesture)
Data: xd
Models:
  xd.null: competitorAdvantage ~ 1 + (1 + gesture + speech
    ↪ | Participant) + xd.null:      (1 | MediaName)
  xd.gesture: competitorAdvantage ~ gesture + (1 + gesture +
    ↪ speech | Participant) + xd.gesture:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
xd.null      9 86560 86641 -43271      86542
xd.gesture  10 86540 86630 -43260      86520 22.297      1
  ↪ 2.336e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

```



```

anova(xd.mannerSub, xd.mannerFull)
Data: xd
Models:
  xd.mannerSub: competitorAdvantage ~ speech + gesture + (1
    ↪ + gesture + speech | xd.mannerSub: Participant)
    ↪ + (1 | MediaName)
  xd.mannerFull: competitorAdvantage ~ speech * gesture + (1
    ↪ + gesture + speech | xd.mannerFull: Participant)
    ↪ + (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
xd.mannerSub  11 86539 86638 -43258      86517
xd.mannerFull 12 86538 86647 -43257      86514 2.4534      1
    ↪      0.1173

lsmeans(xd.mannerFull, pairwise~speech*gesture, adjust="
  ↪ tukey")
$lsmeans
speech gesture      lsmean      SE df  asymp.LCL asymp.
  ↪ UCL
dotted absent  0.24365658 0.02074780 NA 0.20299164
  ↪ 0.2843215
wiggly absent  0.18054049 0.02170494 NA 0.13799960
  ↪ 0.2230814
dotted present 0.10322541 0.01881863 NA 0.06634157
  ↪ 0.1401092
wiggly present 0.09144242 0.02116832 NA 0.04995328
  ↪ 0.1329316

Confidence level used: 0.95

$contrasts
contrast                                estimate      SE df z.
  ↪ ratio p.value
dotted,absent - wiggly,absent  0.06311608 0.02663517 NA
  ↪ 2.370  0.0830
dotted,absent - dotted,present 0.14043117 0.02749728 NA
  ↪ 5.107  <.0001
dotted,absent - wiggly,present 0.15221415 0.03130530 NA
  ↪ 4.862  <.0001
wiggly,absent - dotted,present 0.07731509 0.02954478 NA
  ↪ 2.617  0.0440
wiggly,absent - wiggly,present 0.08909807 0.02750650 NA
  ↪ 3.239  0.0066
dotted,present - wiggly,present 0.01178298 0.02638909 NA
  ↪ 0.447  0.9703

P value adjustment: tukey method for comparing a family of
  ↪ 4 estimates

summary(xd.mannerFull)

```



```

Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: competitorAdvantage ~ speech * gesture + (1 +
  ↳ gesture + speech | Participant) + (1 | MediaName
  ↳ )
Data: xd

AIC      BIC    logLik deviance df.resid
86538.1  86646.8 -43257.1  86514.1    63068

Scaled residuals:
  Min      1Q  Median      3Q      Max
-3.4953 -0.3923 -0.1771  0.0840  2.3993

Random effects:
  Groups      Name              Variance Std.Dev. Corr
MediaName (Intercept)    0.009948 0.09974
Participant (Intercept) 0.005526 0.07434
gesturepresent 0.007670 0.08758 -0.78
speechwiggly 0.005816 0.07626 -0.39 0.14
Residual              0.228122 0.47762
Number of obs: 63080, groups: MediaName, 158; Participant,
  ↳ 35

Fixed effects:
  Estimate Std. Error t value
(Intercept)          0.24366    0.02075  11.744
speechwiggly         -0.06312    0.02664  -2.370
gesturepresent       -0.14043    0.02750  -5.107
speechwiggly:gesturepresent 0.05133    0.03264   1.573

Correlation of Fixed Effects:
  (Intr) spchwg gstrpr
speechwggly -0.605
gesturprnt -0.730  0.405
spchwggly:g  0.398 -0.620 -0.593

#7350-7650
yd <- subset(df3, Timebin >= 7350 & Timebin <= 7650)

yd.null <- lmer(competitorAdvantage ~ 1 + (1+speech+gesture
  ↳ |Participant) + (1|MediaName), data = yd, REML=FALSE)
yd.speech <- lmer(competitorAdvantage ~ speech + (1+speech+
  ↳ gesture|Participant) + (1|MediaName), data = yd, REML
  ↳ =FALSE)
yd.gesture <- lmer(competitorAdvantage ~ gesture + (1+
  ↳ speech+gesture|Participant) + (1|MediaName), data =
  ↳ yd, REML=FALSE)
yd.mannerFull <- lmer(competitorAdvantage ~ speech*gesture
  ↳ + (1+speech+gesture|Participant) + (1|MediaName),
  ↳ data = yd, REML=FALSE)

```

```

yd.mannerSub <- lmer(competitorAdvantage ~ speech + gesture
  ↪ + (1+speech+gesture|Participant) + (1|MediaName),
  ↪ data = yd, REML=FALSE)

anova(yd.null, yd.speech)
Data: yd
Models:
  yd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
    ↪ | Participant) + yd.null:      (1 | MediaName)
  yd.speech: competitorAdvantage ~ speech + (1 + speech +
    ↪ gesture | Participant) + yd.speech:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
yd.null      9 65291 65372 -32636      65273
yd.speech  10 65292 65382 -32636      65272 0.5014      1
  ↪ 0.4789

anova(yd.null, yd.gesture)
Data: yd
Models:
  yd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
    ↪ | Participant) + yd.null:      (1 | MediaName)
  yd.gesture: competitorAdvantage ~ gesture + (1 + speech +
    ↪ gesture | Participant) + yd.gesture:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
yd.null      9 65291 65372 -32636      65273
yd.gesture  10 65285 65376 -32633      65265 7.245      1
  ↪ 0.00711 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1

anova(yd.mannerSub, yd.mannerFull)
Data: yd
Models:
  yd.mannerSub: competitorAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | yd.mannerSub:      Participant
    ↪ ) + (1 | MediaName)
  yd.mannerFull: competitorAdvantage ~ speech * gesture + (1
    ↪ + speech + gesture | yd.mannerFull:      Participant)
    ↪ + (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
yd.mannerSub  11 65287 65387 -32633      65265
yd.mannerFull 12 65289 65398 -32633      65265 5e-04      1
  ↪ 0.9818

lsmeans(yd.mannerFull, pairwise~speech*gesture, adjust="
  ↪ tukey")
$lsmeans

```

```
speech gesture      lsmean          SE df  asymp.LCL  asymp.
  ↳ UCL
dotted absent  0.09452689 0.01668759 NA 0.06181981
  ↳ 0.12723396
wiggly absent  0.09870004 0.01500872 NA 0.06928349
  ↳ 0.12811659
dotted present 0.04309560 0.01474294 NA 0.01419998
  ↳ 0.07199123
wiggly present 0.04779029 0.01560471 NA 0.01720562
  ↳ 0.07837496
```

Confidence level used: 0.95

**\$contrasts**

```
contrast              estimate          SE df
  ↳ z.ratio p.value
dotted,absent - wiggly,absent  -0.004173153 0.01834570 NA
  ↳ -0.227 0.9959
dotted,absent - dotted,present  0.051431281 0.02229786 NA
  ↳ 2.307 0.0965
dotted,absent - wiggly,present  0.046736596 0.02548963 NA
  ↳ 1.834 0.2575
wiggly,absent - dotted,present  0.055604434 0.02215076 NA
  ↳ 2.510 0.0583
wiggly,absent - wiggly,present  0.050909749 0.02230237 NA
  ↳ 2.283 0.1020
dotted,present - wiggly,present -0.004694685 0.01817418 NA
  ↳ -0.258 0.9940
```

P value adjustment: tukey method **for** comparing a **family** of  
 ↳ 4 estimates

**summary**(yd.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: competitorAdvantage ~ speech \* gesture + (1 +  
 ↳ speech + gesture | Participant) + (1 | MediaName)  
 Data: yd

```
AIC      BIC    logLik deviance df.resid
65289.2  65397.7 -32632.6  65265.2    62372
```

Scaled **residuals**:

```
Min      1Q  Median      3Q      Max
-3.5044 -0.2747 -0.1240  0.0318  2.8841
```

Random **effects**:

```
Groups      Name          Variance Std.Dev. Corr
MediaName  (Intercept)  0.004725 0.06874
Participant (Intercept)  0.005049 0.07105
speechwiggly 0.002467 0.04967 -0.61
```

```
gesturepresent 0.008164 0.09036 -0.79 0.31
Residual 0.164948 0.40614
Number of obs: 62384, groups: MediaName, 158; Participant,
↳ 35
```

Fixed **effects**:

	Estimate	Std. Error	t	value
(Intercept)	0.0945269	0.0166876	5.665	
speechwiggly	0.0041732	0.0183457	0.227	
gesturepresent	-0.0514313	0.0222979	-2.307	
speechwiggly:gesturepresent	0.0005215	0.0228271	0.023	

Correlation of Fixed Effects:

```
(Intr) spchwg gstrpr
speechwggly -0.637
gesturprsnt -0.750 0.419
spchwggly:g 0.346 -0.630 -0.512
```

#7700-7950

```
zd <- subset(df3, Timebin >= 7650 & Timebin <= 7950)
```

```
zd.null <- lmer(competitorAdvantage ~ 1 + (1+speech+gesture
↳ |Participant) + (1|MediaName), data = zd, REML=FALSE)
```

```
zd.speech <- lmer(competitorAdvantage ~ speech + (1+speech+
↳ gesture|Participant) + (1|MediaName), data = zd, REML
↳ =FALSE)
```

```
zd.gesture <- lmer(competitorAdvantage ~ gesture + (1+
↳ speech+gesture|Participant) + (1|MediaName), data =
↳ zd, REML=FALSE)
```

```
zd.mannerFull <- lmer(competitorAdvantage ~ speech*gesture
↳ + (1+speech+gesture|Participant) + (1|MediaName),
↳ data = zd, REML=FALSE)
```

```
zd.mannerSub <- lmer(competitorAdvantage ~ speech + gesture
↳ + (1+speech+gesture|Participant) + (1|MediaName),
↳ data = zd, REML=FALSE)
```

```
anova(zd.null, zd.gesture)
```

Data: zd

Models:

```
zd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
↳ | Participant) + zd.null: (1 | MediaName)
zd.gesture: competitorAdvantage ~ gesture + (1 + speech +
↳ gesture | Participant) + zd.gesture: (1 |
↳ MediaName)
```

Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
zd.null	9	37447	37528	-18714	37429			
zd.gesture	10	37449	37539	-18714	37429	4e-04	1	↳ 0.9834

```
anova(zd.null, zd.speech)
```

```

Data: zd
Models:
  zd.null: competitorAdvantage ~ 1 + (1 + speech + gesture
    ↪ | Participant) + zd.null:      (1 | MediaName)
  zd.speech: competitorAdvantage ~ speech + (1 + speech +
    ↪ gesture | Participant) + zd.speech:      (1 |
    ↪ MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
zd.null      9 37447 37528 -18714      37429
zd.speech  10 37448 37538 -18714      37428 0.7273      1
    ↪ 0.3937

anova(zd.mannerSub, zd.mannerFull)
Data: zd
Models:
  zd.mannerSub: competitorAdvantage ~ speech + gesture + (1
    ↪ + speech + gesture | zd.mannerSub:      Participant
    ↪ ) + (1 | MediaName)
  zd.mannerFull: competitorAdvantage ~ speech * gesture + (1
    ↪ + speech + gesture | zd.mannerFull:      Participant)
    ↪ + (1 | MediaName)
Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
zd.mannerSub  11 37450 37549 -18714      37428
zd.mannerFull 12 37452 37560 -18714      37428      0      1
    ↪          0.998

lsmeans(zd.mannerFull, pairwise~speech*gesture, adjust="
    ↪ tukey")
$lsmeans
speech gesture      lsmean          SE df  asymp.LCL  asymp.
    ↪ UCL
dotted absent  0.05818611 0.01305903 NA 0.03259088
    ↪ 0.08378134
wiggly absent  0.04658453 0.01397805 NA 0.01918806
    ↪ 0.07398101
dotted present 0.05739782 0.01344709 NA 0.03104200
    ↪ 0.08375363
wiggly present 0.04584568 0.01487034 NA 0.01670034
    ↪ 0.07499102

Confidence level used: 0.95

$contrasts
contrast                estimate          SE df
    ↪ z.ratio p.value
dotted,absent - wiggly,absent  0.0116015761 0.01677466 NA
    ↪ 0.692 0.9003
dotted,absent - dotted,present 0.0007882929 0.01674322 NA
    ↪ 0.047 1.0000
dotted,absent - wiggly,present 0.0123404282 0.01948914 NA

```

```

↪      0.633  0.9214
wiggly,absent - dotted,present  -0.0108132832  0.01867021 NA
↪      -0.579  0.9384
wiggly,absent - wiggly,present   0.0007388521  0.01674314 NA
↪      0.044  1.0000
dotted,present - wiggly,present  0.0115521353  0.01663358 NA
↪      0.695  0.8992

```

P value adjustment: tukey method **for** comparing a **family** of  
 ↪ 4 estimates

**summary**(zd.mannerFull)

Linear mixed **model** fit **by** maximum likelihood ['lmerMod']  
 Formula: competitorAdvantage ~ speech \* gesture + (1 +  
 ↪ speech + gesture | Participant) + (1 | MediaName  
 ↪ )

Data: zd

AIC	BIC	logLik	<b>deviance</b>	<b>df.resid</b>
37452.3	37560.5	-18714.2	37428.3	60638

Scaled **residuals**:

Min	1Q	Median	3Q	Max
-3.9015	-0.2601	-0.1067	0.0478	3.6341

Random **effects**:

Groups	Name	Variance	Std.Dev.	Corr
MediaName	(Intercept)	0.003572	0.05977	
Participant	(Intercept)	0.002461	0.04961	
speechwiggly	0.002866	0.05353	-0.38	
gesturepresent	0.002911	0.05395	-0.46	0.10
Residual		0.107294	0.32756	

Number of obs: 60650, groups: MediaName, 158; Participant,  
 ↪ 35

Fixed **effects**:

Estimate	Std. Error	t	value
(Intercept)		5.819e-02	1.306e-02 4.456
speechwiggly		-1.160e-02	1.677e-02 -0.692
gesturepresent		-7.883e-04	1.674e-02 -0.047
speechwiggly:gesturepresent		4.944e-05	1.976e-02 0.003

Correlation of Fixed Effects:

(Intr)	spchwg	gstrpr
speechwggly	-0.586	
gesturprsnt	-0.618	0.379
spchwggly:g	0.382	-0.596 -0.590