THE UNIVERSITY OF CHICAGO

A COMMUNICATIVE ACCOUNT OF GESTURING WHEN SPEAKING IS DIFFICULT

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY
YAĞMUR DENIZ KISA

CHICAGO, ILLINOIS
JUNE 2022

Annem ve babama

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

About five years ago I moved to the other side of the world to learn a craft from the best, a dream of mine — sadly leaving behind my family, friends, and the city I grew up in. Throughout these years, I traveled back and forth between Chicago, Ithaca, Brussels, and Istanbul. I am so grateful for all the loved ones who supported me during this transatlantic journey and for all the amazing people I got to know thanks to this journey.

Susan Goldin-Meadow & Daniel Casasanto: Thank you so much for giving me the opportunity to learn how to be a scientist from you. Your love for big questions and for thinking together with people gave so much meaning to doing science.

Kensy Cooperrider, Geoffrey Brookshire, Defu Yap, Benjamin Pitt, Amritpal Singh, Emma Murrugarra, Haily Merritt, Jennifer Lu, James Waller & Casey Ferrara: I am so grateful that I got to have these brilliant and kind people as my colleagues throughout the years.

All the research assistants I worked with, especially Cordelia Achen: Thank you for the many hours of hard work.

Addie Barron: Thank you for being the beautiful person you are and thanks for making Hotel Feelings such a happy and safe place to be during a tough year.

Hannah Kim: Thank you for being there for me from day one and waiting with open arms every time I came back to one of our homes.

Cameron Knox Day & Miso: Thank you for being such sweet housemates.

Joseph Gogola: Thank you for being my amazing husband, for your big heart, and for all the hugs and kisses.

Benjamin Morris: Thank you for all the chitchats, plays, and hugs. I am so lucky to have such a sweet soul as a friend.

Claire Bergey: We did this together and I am sad that we don't get to say goodbye together. Thank you for ending up in the same school with me – being here was much more

meaningful thanks to you and your beautiful mind.

Celine & Joe Cammarata: Thank you for giving me a room of my own in Ithaca — you guys were the sweetest family to live with.

Fred De Ryck, Sofie Robberechts, Pieter Gijssels & Alice Van Innis: Thank you for opening the doors of your homes to me, making me feel always so welcome and loved in Brussels, and all the amazing babies you made over the years that I get to hug and play with.

Saniye Ince: Canım kankam. Varlığın varlığıma armağan. Beni gurbet ellerde hiç yalnız koymadın. Hayatımda olduğun için çok şanslıyım.

Alper Şükrü Gençer: Canım hocam, bu hayali beraber kurduk, beraberinde gelen iyileri de kötüleri de beraber yaşadık. Bu yolculuğu her adımında seninle paylaştığım için çok şanslıyım, güzel yürekli arkadaşım.

Ece Zeynep Taşkın: Canım arkadaşım, seni bırakıp uzaklara gittiğim için özür dilerim. Bana olan her daim, koşulsuz sevgin için sana çok teşekkür ederim.

My mom, Elif Aslan: Anacım, bana sonsuz inancın ve tüm emeklerin için teşekkür ederim.

My dad, Bektaş Kısa: Babacım, hayallerimin hep en büyük destekçisi olduğun için teşekkür ederim. Bana hayatı ve öğrenmeyi sevmeyi öğrettiğin için de...

Tom Gijssels: Thank you for being at the same corner of the world, at the same time, because of the same passion — so that I got to find you. Thank you for being the curious, playful, emotional and brilliant person you are. Thank you for trusting us, always. Thank you for being patient with making your dreams come true, so that I could work towards mine. I'm done my love, finally, I'm coming home.

ABSTRACT

People gesture when they speak, especially when speaking is difficult. People are more likely to gesture when their speech is disfluent or when they are about to utter a low probability word. Why? According to an influential answer to this question, people gesture when speaking is difficult because gestures facilitate speech production – serving a cognitive function for the speaker. In Study 1, I aimed to conceptually replicate the finding that preventing gestures makes speakers more disfluent – the primary finding that has been interpreted as evidence for this influential hypothesis. Contrary to long-held beliefs, I showed that there is, in fact, no support for the claim that preventing gestures impairs speaking – in a large dataset and re-examining five decades of empirical evidence cited for this claim. If gestures do not help speech production, why then are people more likely to gesture when speaking is difficult? To answer this question, I proposed and tested an alternative explanation. Gesturing when speaking is difficult may not be a mere symptom of speaking difficulties. And gesturing may not be produced to help speakers resolve these difficulties, either. Instead, gesturing when speaking is difficult may be communicatively motivated. In Study 2, I showed that speakers are more likely to gesture when they are disfluent because gestures serve as a pragmatic signal to the listener commenting on experiencing problems with speaking. When gestures were not visible and therefore could not serve as a pragmatic signal to the listener, they were not more likely to occur during disfluent speech than fluent speech. In Study 3, I showed that speakers are more likely to gesture before they are about to utter a low probability word because gestures signal and depict hard to predict words for the listener. When gestures were not visible and could not serve as a communicative signal to the listener, they were not more likely to occur before uttering a low probability word than a high probability word. Together, these three studies motivate a revision to theories of why people gesture. People gesture when speaking is difficult, because gestures provide a communicative signal to the listener, commenting on the process of speaking or the content of upcoming words.

CHAPTER 1

INTRODUCTION

When people speak, they gesture. Why? Some gestures clearly serve a communicative function. For example, people use a “thumbs up” gesture, an emblem, to signal approval or point to a cup to indicate where the cup is. In cases like these, gesture is an essential part of the communicative act: If somebody were to say “there” without pointing, the sentence would be incomplete.

But for other gestures, many researchers claimed that it is not so clear why people produce them (for example see Krauss et al., 1991). The gestures people produce when they speak are not limited to emblems, such as “thumbs up,” or points. People also use other gestures: Depicting ideas with iconics or metaphorics and producing beats in line with the rhythm of speech. Iconics depict concrete things and/or actions such as making a holding shape to depict holding a cup. Metaphorics depict abstract things as if they are concrete things and/or actions (e.g. metaphorically holding ideas in hands when one says “on the one hand”). Beats, in terms of their form, seem like mere flicks of the hand, beating along with the rhythm of speech (but see Yap et al. 2018 for beats that represent speech). These gestures have been argued to be more easily dispensable, that is, not essential to the communicative message: Even if people did not use gestures to depict or beat, their sentences would be considered complete.

Why do people produce these gestures then? What is their function, which leads to their recurrence? Many researchers argued that gestures, even ones that may seem not essential to the communicative message, nevertheless serve clear communicative functions for the listener (Bavelas et al., 1992; Clark, 1996; Holler & Levinson, 2019; Kendon, 2004; McNeill, 2016). Listeners pay attention to gestures and change their understanding of speakers’ communicative message using the semantic information conveyed in gestures (see for example Alibali et al., 1997; Holler et al., 2009). Gestures effectively communicate not only semantic information, but also pragmatic information. For example, when speakers gesture to seek agreement, listeners respond accordingly by providing agreement; when speakers gesture to hand over their turn, the speakers respond accordingly by taking the turn (Bavelas et al., 1995). So, gestures effectively communicate ideas to the listeners.

Many others argued that gesture's functions are not limited to communicative functions for the listener and that gestures also serve cognitive functions for the speaker (Butterworth & Hadar, 1989; Goldin-Meadow, 1999; Kita et al. 2017; Krauss, 1998). One of the first functional explanations for a speaker-internal cognitive function of gesturing proposed that gestures help people speak (Butterworth & Hadar, 1989; Krauss & Hadar, 1999; Rauscher et al. 1996). According to an influential version of this proposal, the Lexical Retrieval Hypothesis (LRH), gestures facilitate speech production by helping speakers find the right words. Typically, reviews listing the cognitive functions of gesture start this list with the assertion that gesture facilitates speech production (e.g., Alibali, 2005; Goldin-Meadow, 1999; Kita et al., 2017).

The LRH seemed like a very plausible hypothesis, partly, because of a very robust relationship between speech and gesture: People are more likely to gesture when speaking is difficult. People are particularly likely to gesture when they cannot find the right word, when their speech is not fluent, or when they produce low probability words or less preferred sentences (Beattie & Aboudan, 1994; Beattie & Shovelton, 2000; Cook et al., 2009; Frick-Horbury & Guttentag, 1998; Nobe, 2000; Pine et al., 2007).

The LRH offered a potential functional explanation for why people might resort to gesturing when speaking is difficult: Because gestures may help speaking. Even the proponents of the LRH conceded that some gestures serve functions for the listener (see Krauss et al., 1991), but gestures that occur when speaking is difficult have been interpreted as evidence that they must be for the speaker, helping their own cognitive functions, rather than for the listener (see for example Krauss & Hadar, 1999).

What is the evidence for the claim that gestures serve a cognitive function for the speaker by helping them speak? To test whether gestures serve a function for the speaker, one must manipulate gestures and see if it has any consequences on speaking. One study by Krauss and colleagues (Rauscher et al., 1996) prevented people from gesturing to see if it hurts speaking. According to Krauss and colleagues (Rauscher et al. 1996), if people have difficulty finding words, their speech should be more disfluent. If gesturing helps speakers find the right words, then preventing speakers from gesturing should make their speech production more disfluent, compared to when they are free to gesture. In an influential study, Krauss and colleagues claimed that preventing people from gesturing did indeed make their speech more disfluent (Rauscher et al., 1996). This study by Krauss and colleagues (Rauscher et al. 1996) has been widely cited as evidence for the idea that gesturing helps speakers find the right words and, more broadly, as some of the first evidence that gesturing serves a cognitive function for speakers. Largely based on this one study, the field has come to accept the hypothesis that people gesture, in part, because gestures facilitate speech production.

In this study, Krauss and colleagues (1996) claimed gestures may not help just any kind of speech, but particularly spatial speech. They found that people gesture far more frequently during speech with spatial content rather than speech without spatial content and claimed that gesture prevention selectively makes spatial speech disfluent. In Chapter 2 (Study 1), I sought to conceptually replicate Rauscher et al.'s (1996) finding that preventing speakers from gesturing increases disfluencies in their speech with literal spatial content (e.g., the rocket went up). I also aimed to extend this pattern to speech with metaphorical spatial content (e.g., my grades went up) to test whether gestures also help speakers find the right words when they talk about abstract concepts that are spatialized metaphorically. Much to my surprise, I found no difference in disfluency between speakers who were allowed to gesture freely and speakers who were not allowed to gesture, for any of the measures Rauscher et al. used, in any category of speech (literal spatial content, metaphorical spatial content, no spatial content). In a large dataset (7969 spoken phrases containing 2075 disfluencies), I found no support for the idea that gestures help speakers find the right words, even for speech with literal spatial content.

Failing to conceptually replicate Rauscher et al.'s (1996) influential study motivated a detailed reexamination of all of the work that gets cited as evidence for the LRH – including Rauscher et al.'s (1996) study. Is there any evidence that gesture prevention hurts speech production? Upon reexamining Rauscher et al.'s (1996) claim that gesture prevention makes spatial speech disfluent, I found that their own data do not support this influential claim. My further reexamination of five decades of research testing versions of this hypothesis, before and after Rauscher et al.'s (1996) influential study, also revealed that there is no reliable evidence that gesture prevention hurts speech production. Based on this detailed re-examination of past work, I conclude that gestures do not appear to facilitate speech production, challenging long-held beliefs about why people gesture when they speak.

Given that gestures do not appear to help speaking, why then are people more likely to gesture when speaking is difficult? The LRH offered both a mechanistic and a functional account of why people gesture, particularly when speaking is difficult. A mechanistic explanation for why people gesture would ask about the preceding events that cause gesturing to happen. On the other hand, a functionalist explanation for why people gesture, would ask about the consequences of gesturing that leads to its recurrence. Krauss and colleagues proposed both a mechanism for gesturing, that speech difficulties cause gestures, and a function of gesturing, that these gestures help resolve speech difficulties. Given that resolving speech difficulties does not seem to be a function of gesturing, are gestures produced because they are mere symptoms of speech difficulties? That is, does the LRH, nevertheless, offer a good mechanistic explanation for why people gesture? In principle, it is possible that speech difficulties are a reliable mechanism that elicits gesturing – but they are not functional with respect to resolving those speech difficulties.

Alternatively, speech difficulties may co-occur with gestures, because those gestures are communicatively motivated. People may gesture when speaking is difficult, not because gestures are mere symptoms of speech difficulties, but because gestures aim to facilitate the conversation by helping the listener. I tested this mechanistic explanation for gesturing when speaking is difficult in chapters 3 (Study 2) and 4 (Study 3).

1.1 A pragmatic account of why people gesture during disfluent speech

People are more likely to gesture when their speech is disfluent, compared to when their speech is fluent (Akhavan et al., 2016; Butterworth & Beattie, 1978; Ragsdale & Silvia, 1982). This co-occurrence of gestures with disfluencies motivated the idea that speaking difficulties cause gesturing (a speaker-internal cognitive mechanism) and gestures help people resolve speech disfluencies by helping them find the right thing to say (a speaker-internal cognitive function). In Chapter 3 (Study 2), I tested whether gesturing during disfluent speech might instead be pragmatically motivated.

People may gesture when their speech is less fluent because gestures serve as a pragmatic signal to the listener, commenting on experiencing problems with speaking. Gestures can comment in many ways. For example, speakers can foreshadow oncoming interruption or signal the speaker's commitment to a fluent re-start. In these and many other ways, gesturing during disfluent speech may give the listener information about the speaker's production plan. This information, in turn, can ensure successful coordination in conversation timing (Holler & Levinson, 2019) and also prevent interpersonal damage from not meeting the social expectations in conversing: Conversation is a joint activity and speakers would be violating the principle of being cooperative if they do not acknowledge their deviation from the role they commit to as a speaker (presenting an utterance, Clark, 1996).

If gesturing during disfluent speech is pragmatically motivated, when the listener cannot see the speaker, the speaker's pragmatic motivation to gesture during disfluencies should disappear. Alternatively, if gesturing during disfluent speech is not pragmatically motivated, but merely is a symptom of speech difficulties, then the visibility of gestures should not matter: People should be likely to gesture more during disfluent speech regardless of whether their gestures are visible or not. In Study 2, I show evidence supporting the pragmatic account: Speakers were more likely to gesture during disfluent speech only when the listener could see their gestures, not when the listener was prevented from seeing their gestures.

1.2 A communicative account of why people gesture to depict low probability words

In Chapter 4, I turn to another case of gesturing when speaking is difficult. Semantic gestures tend to depict words that have a low probability of continuing that utterance (Beattie & Shovelton, 2000) – another case when speaking is difficult. Low probability words are words that are likely harder to retrieve from the mental lexicon for the speaker. This link between gestures and 'hard to retrieve' words motivated the idea that retrieval difficulties cause gestures and producing gestures help the speaker retrieve these words more easily (Hadar & Butterworth, 1997; Krauss & Hadar, 1999). In Chapter 4 (Study 3), I tested whether this link between gestures and low probability words is communicatively motivated instead.

Speakers may be more likely to gesture to depict words that have a low probability of continuing that utterance because gestures are produced with a communicative motivation to signal and depict unpredictable words for the listener. Speakers may be more likely to gesture before a low probability word than a high probability word because the gestures may serve as a pragmatic signal to the listener that a relatively less predictable word is coming up. Additionally, the form of the gesture may depict information for the listener, so that they can more easily predict and process relatively unpredictable, highly informative, parts of the utterance.

If gesturing to depict less predictable words is indeed communicatively motivated, then when the listener cannot see the speaker, the speaker's communicative motivation to depict words with a lower transitional probability in gestures should disappear or weaken. Alternatively, if gesturing to depict less predictable words is a symptom of word finding difficulties, gestures' lexical affiliates should have a lower transitional probability compared to other words to the same extent when the listener can and cannot see the gestures. In Study 3, I show evidence supporting the communicative account: Speakers gestured to depict lower probability words only when the listener could see their gestures, not when the listener was prevented from seeing their gestures.

Together, this dissertation motivates a major revision to theories of why people gesture – about both the functions and the mechanisms of gesturing. Contrary to long-held beliefs, I show that helping speech production is not one of the cognitive functions of gesturing. Gesture's function is not limited to communicative ones, but in light of the evidence I present here, an account of the cognitive functions of gesture will need to look beyond the role of gesture in speech production. I also argue that gestures are not mere symptoms of speech difficulties and propose an alternative mechanism for why people gesture, especially when speaking is difficult. I show that people may gesture when they are disfluent or when they are about to say a difficult word, not because these gestures grow out of speaking difficulties or they help speaking, but because they aim to facilitate the conversation by helping the listener.

CHAPTER 2

DO GESTURES REALLY FACILITATE SPEECH PRODUCTION?

2.1 Introduction

Why do we gesture when we speak? For some gestures, it is self-evident that the speaker intends them to be communicative, as when we wave "hello" or give a "thumbs up" gesture to signal approval. Do speakers also gesture because gesturing serves a cognitive function in the speaker's mind, helping them to think or to talk?

For decades, researchers have posited that gestures facilitate speech production (Butterworth & Hadar, 1989; Krauss, 1998; Krauss & Hadar, 1999). According to an influential version of this proposal, the Lexical Retrieval Hypothesis (LRH), gestures facilitate speech production by helping speakers find the right words; however, only *some* gestures are posited to affect some words. Krauss and colleagues (Krauss, 1998; Rauscher et al., 1996) noted that people gesture far more frequently during phrases with spatial content than during phrases without it, and hypothesized that *gestures that reflect spatial features of meaning* helps speakers find the right *spatial words*. Krauss and colleagues hypothesized that some gestures "derive from knowledge encoded in a spatial format" (Rauscher et al., 1996, p. 227); the spatial features of a gesture (e.g., upward trajectory) facilitate production of words by priming the spatial features (e.g., upwardness) that enter into the search for that word (e.g., for the word "up"; Krauss & Hadar, 1999).

What is the evidence that gesturing helps speakers find the right spatial words? According to Krauss and colleagues (Rauscher et al., 1996), if people have difficulty finding words, their speech should be more disfluent. If gesturing helps speakers find the right words, then preventing speakers from gesturing should make their speech production more disfluent, compared with when they are free to gesture. In an influential study, Krauss and colleagues reported that preventing people from gesturing increased the number of disfluencies they produced (pauses, repairs, etc.), and slowed down their speech selectively for the production of spatial phrases (Rauscher et al., 1996). Additionally, preventing people from gesturing when they produced spatial clauses increased their rate of filled pauses within those clauses (nonjuncture filled pauses), relative to filled pauses at the junctures of the clauses (i.e., juncture filled pauses), which Rauscher and colleagues claimed is the measure that most sensitively reflects problems in word finding. This study is frequently cited as evidence that gesturing helps speakers find the right spatial words and is discussed in reviews summarizing the state of knowledge about speaker-internal functions of gesture (e.g., Goldin-Meadow, 1999; Hostetter & Alibali, 2008). Largely based on this one study, the field has come to accept that one speaker-internal cognitive function played by gesture is to help speakers find the right words and facilitate speech production (e.g., Alibali et al., 2000, 2011; Casasanto, 2013; Goldin-Meadow, 1999; Hoetjes et al., 2014; Hostetter, 2011; Iverson & Goldin-Meadow, 1998; Krauss, 1998).

If LRH only explains how gestures help people talk about *space*, however, it provides only a limited account of how gesturing helps speaking. People spend a lot of time speaking about nonspatial ideas, including highly abstract concepts: entities like *time* and *value* that have no spatial magnitude, direction, or location. Yet, there is abundant evidence that people use space metaphorically to speak, think, and gesture about abstract concepts (Cienki, 2005; Lakoff & Johnson, 1980; McNeill, 1992; for a review, see Casasanto & Bottini, 2014). In one study, on which the present study builds, people spontaneously produced gestures whose form reflected the spatial direction implied in their speech (e.g., upward), regardless of whether they talked about concrete space (e.g., “the rocket went up”) or metaphoric space (e.g., “my grades went up”; Yap et al., 2018). People’s gestures reflected the predicted spatial directions (e.g., better is metaphorically upward), even when they talked about abstract concepts without using any spatial words (e.g., “my grades got better”). According to this study, which analyzed over 5,000 gestures, people were just as likely to gesture spontaneously in the predicted directions for metaphorically spatialized concepts (e.g., grades are not the kind of entity that can literally rise in space) as for literal spatial concepts. These results suggest that, like words for literal spatial concepts, words for metaphorical spatial concepts correspond to particular kinds of spatial information in speakers' minds (e.g., schematic representations of upward, downward, rightward, or leftward space), even when these words have no literal spatial uses (e.g., the word "better" cannot be used sensibly to denote literal spatial locations or paths).

If spatial information is activated in memory not only when people produce literal spatial language, but also when they produce metaphorical spatial language, then the same gestural mechanism should help people find words for both literal spatial scenarios and metaphorically spatialized ideas. Gesturing upward, for example, should help speakers not only to produce words or phrases like "my rocket went up," but also to produce words or phrases like "my grades went up" and perhaps even "my grades got better." If so, this discovery would substantially expand the scope of gesture's role in speech production, particularly since spatial schemas appear to be part of people's mental representations in many nonspatial conceptual domains that become spatialized metaphorically in language and thought, including time (Clark, 1973b), number (Dehaene et al., 1993), emotional valence (Casasanto, 2009), power (Schubert, 2005), similarity (Casasanto, 2008), intimacy (Matthews & Matlock, 2011), and musical pitch (Rusconi et al., 2006), among others (Lakoff & Johnson, 1980).

Here, we tested whether gestures serve a speaker-internal cognitive function by helping people find the right words with literal or metaphorical spatial content. We sought first to conceptually replicate Rauscher and colleagues' (Rauscher et al., 1996) study testing whether gesturing helps speakers produce words for literal spatial scenarios, and then to determine whether this benefit extends to producing speech about metaphorically spatialized ideas. We compared how fluently people spoke when they were allowed to gesture freely and when they were prevented from gesturing as they told stories with either literal or metaphorical spatial content. To measure disfluency in speech production, we calculated the number of speech disfluencies (repair, repeat, etc.) per word, speech rate (number of words per minute), and the relative frequency of nonjuncture filled pauses, following Rauscher et al. (1996). If gesturing only helps people find the right concrete spatial words, as suggested by Rauscher and colleagues (Rauscher et al., 1996), then preventing people from gesturing should make speech production more disfluent (i.e., higher disfluency rate, slower speech rate, and more nonjuncture than juncture filled pauses) only for speech with literal spatial content. Alternatively, if gesturing can also help speakers find the right abstract words, then preventing people from gesturing should make speech production more disfluent for speech not only with literal spatial content, but also with metaphorical spatial content.

Several studies in the 5 decades before and after Rauscher et al. (1996) have also tested whether preventing speakers from gesturing make their speech production more disfluent and reported null effects (Cravotta et al., 2018; Finlayson et al., 2003; Graham & Heywood, 1975; Hoetjes et al., 2014; Hostetter et al, 2007; Rimé et al., 1984). However, these null effects could simply be because none of the studies distinguished effects of gesture prevention on spatial and nonspatial speech; thus, arguably, these studies did not attempt to conceptually replicate Rauscher et al.'s (1996) claim that gesture prevention selectively affects spatial speech. Here, we tested this claim explicitly.

To preview our findings, contrary to our expectations based on earlier claims, preventing gesture had no significant effect on any of our planned dependent measures—not for speech with metaphorical spatial content, and not even for speech with literal spatial content. Due to the large number of data points in our study, it is not likely that the absence of these effects was the result of low statistical power. In response to this unexpected outcome, we first scrutinized our own data to confirm that there was no effect of gesture prevention *beyond* our planned analyses, in any subsets of the data (i.e., disfluency rates for different types of disfluencies). We then scrutinized the results of previous studies testing the LRH and related proposals, over the past 5 decades. Upon reexamining these studies, we conclude that (as in the present study), preventing gesture had no interpretable effect on speech production, even for literal spatial words, motivating a reexamination of widely held beliefs about why people gesture when they speak.

2.2 Methods

2.2.1 Participants

Fifty-six Stanford University undergraduates (28 male) were recruited in pairs, and participated for course credit after giving informed consent (the study was approved by Stanford University's Institutional Review Board).

2.2.2 Materials

There were 12 brief stories in total, each 50–100 words, implying motion or extension in one of two spatial axes: horizontal or vertical. Four of the stories had literal spatial content,describing actual spatial scenarios in the physical world using concrete spatial directions (e.g., "the rocket went higher"; "the scuba diver went down"). Eight of the stories had metaphorical spatial content, describing abstract nonspatial phenomena that are nevertheless commonly talked and thought about using spatial directions metaphorically (e.g., "my grades went higher"; "the price went down"). Each of the eight metaphorical stories had two versions: metaphorical stories with spatial language and metaphorical stories without spatial language. Metaphorical stories with spatial language described nonspatial phenomena using spatial words or phrases in their abstract metaphoric senses (e.g., "my grades went higher"). Metaphorical stories without spatial language were identical to metaphorical stories with spatial language, except that spatial words or phrases that are used metaphorically were replaced with nonspatial paraphrases conveying nearly the same meaning and implying the same spatial directions (e.g., “my grades got better”). The stories involved an overall spatial direction (horizontal or vertical); however, they contained words and phrases that expressed spatial ideas other than directions or positions such as “long,” “crossing,” “around,” “plunge,” “stretch back,” “boost up,” “stuck at the top,” and so forth.

2.2.3 Procedure

Participants were told that the experiment was about storytelling. They took turns studying written stories, each for 60 seconds, and then retelling the stories to their partners. They were told to retell the stories as accurately as possible because their partner would be quizzed on the content of the stories. All stories were written in the second person (e.g., “You’re testing some new model rockets”), but participants were asked to retell the stories in the first person (e.g., “I’m testing some new model rockets”) as if retelling their own experiences.

After starting with a warm-up story, each participant retold six stories in randomized order: two stories with literal spatial content and four stories with metaphorical spatial content. Each pair of participants received only one version of each metaphorical story: either with spatial language or without spatial language (i.e., one pair of participants would receive either the story about “grades going higher” or the story about “grades getting better”).

Each pair of participants was assigned to one of two gesture conditions: gesture prevented or gesture allowed. In the gesture prevented condition, participants were instructed to hold down keys on a computer keyboard, one key with each hand, during the entire time they were retelling the stories. They were told that the keys activated the microphones mounted on top of the computer monitor in front of them; in fact, the microphones were nonfunctional. In the gesture allowed condition, participants simply told the stories without being instructed to hold down keys on a keyboard; they were not told to gesture. Testing lasted 20–30 min.

2.2.4 Coding

Analyses of the gestures from the Gesture Allowed condition were reported in Yap et al. (2018), but no analysis of speech disfluencies was reported, and no data from the Gesture Prevented condition have been reported previously. In the Gesture Allowed condition, participants produced a total of 2,249 gestures including 1,609 beats, 328 iconic gestures, 252 deictics, 48 metaphoric gestures, 10 adaptors, and two emblems. Beats were categorized solely based on form, following McNeill's (1992) beat filter. When categorized based on meaning with respect to accompanying speech, 629 of the 1,609 beats reflected the spatial ideas expressed in the accompanying speech. Furthermore, the overall rate of gestures that would be predicted to facilitate spatial speech (i.e., beats reflecting spatial semantics, iconic gestures, deictics, and metaphoric gestures) in the Gestures Allowed condition was 56% (1,257 out of 2,249 gestures). We used ELAN (Wittenburg et al., 2006) to code speech disfluencies.

Speech Content Coding

Participants' audio recordings of the stories were transcribed verbatim; 22 of the 336 stories were excluded because the speech was inaudible. The transcriptions of participants' audio recordings of the stories were parsed into clauses and phrases. Coder 1 determined whether each phrase had spatial content, literal or metaphorical. A given phrase was classified as having spatial content if it contained language that implies literal or metaphorical motion, extent, or position along either the lateral or vertical axis. For instance, "went higher" in the rocket story would be a phrase with spatial content since the overall story has a vertical spatial schema of a rocket going higher and higher, and "higher" implies literal motion along the vertical axis. Similarly, "came back down" in the rocket story would be a phrase with spatial content, because "down" also implies literal motion along the vertical axis. Alternatively, phrases were classified as having no spatial content if they did not imply a literal or metaphorical spatial schema.

Spatial Content Type Coding

Phrases with spatial content were classified as literal or metaphorical, and phrases with metaphorical spatial content were further classified as having or not having spatial language, using the same criteria used to construct the stories. Participants produced a total of 7,969 spoken phrases, with very similar numbers of phrases across gesture conditions (Gesture allowed 4,000; Gesture prevented: 3,969). Overall, 2,801 phrases (35% of all phrases) included spatial content, with 962 literal and 1,839 metaphorical spatial content. Coder 2 determined speech content for 10% of all phrases and the intercoder agreement for speech content was 96% (Cohen's $\kappa = .92, z = 26.1, p < .0001$).

Disfluency Rate Coding

Coder 1 recorded the location and type of speech disfluency for each story. Speech disfluencies included repeats, repairs, filled pauses (uh, um, etc.), and unfilled pauses. Coder 2 coded the number of speech disfluencies for a random subset of the stories (29 stories out of a total of 314 stories: about 10%). The intercoder agreement for the number of speech disfluencies for each story was very high; Coder 2's coding explained 94% of the variance in Coder 1's coding (intercoder correlation for number of disfluencies per story: $\beta = .94$, $R^2 = .94, p < .0001$).

Participants produced a total of 2,075 disfluencies (Gesture allowed: 1,041; Gesture prevented: 1,034). Speech disfluencies included 905 (44%) filled pauses, 492 (24%) repairs, 446 (21%) unfilled pauses, and 232 (11%) repeats. For each story, we calculated the total number of disfluencies that occurred during phrases with spatial content (both overall and for each spatial content type separately) and during phrases without spatial content. Finally, for each story, we calculated the total number of words in phrases with and without spatial content, which we used as a baseline in our analysis.

The analyses on disfluency rate were done at the story level, rather than at the phrase level, to reduce the inflated zero problem of our count data (i.e., too many phrases with no disfluencies). When comparing the overall effect of gesture prevention on disfluency rate, we had one observed rate for each story; when comparing the effect of gesture prevention for speech with spatial content versus for speech without spatial content, we had two observed rates for each story; and so on.

Speech Rate Coding

For each phrase, we measured its duration and coded the number of words it contained. We computed the speech rate for each phrase, dividing the number of words by the duration for each phrase. Participants had an average speech rate of 212 words per minute at the phrase level (SD = 86.25). When we computed the speech rate for each story (N = 336), dividing the total number of words by the total duration of all the phrases in each story, participants had an average speech rate of 171 words per minute (SD = 38.86). The analyses were done with speech rates calculated at the phrase level.

Nonjuncture Filled Pause Coding

Each filled pause (905 in total) was classified as nonjuncture or juncture, based on whether it occurred between clause junctions, that is, within a clause (e.g., "And the rocket um shoots up as high as the 10th floor"), or at clause junctures (e.g., "*Um* and the rocket shoots up as high as the 10th floor"). Twelve filled pauses were excluded from the analyses since they occurred during clauses with mixed spatial content type (e.g., both literal and metaphorical spatial content).

2.2.5 Analyses

We conducted all analyses by fitting generalized linear mixedeffect models, using R (R Core Team, 2020), the glmer() function in the lme4 library (Bates et al., 2015), and the optimx package (Nash & Varadhan, 2011). We treated Subject (N = 56) and Story (N = 12) as random effects, including random intercepts for both in analyses, since our outcome variable, disfluency, is likely to vary across different subjects and stories. Subjects and items (i.e., stories here) usually do vary idiosyncratically not only in their global mean responses, but also in their sensitivity to experimental treatment. We used "maximal" random effect structures justified by our design (Barr et al., 2013), including not only random intercepts for Subject and Story, but also random slopes for our fixed factors that are within-subject or within-story, allowing disfluencies of subjects and stories to vary differentially based on our fixed factors (e.g., subjects could be affected differently by speech content manipulation in their production of disfluencies).

Gesture condition (gesture allowed; gesture prevented) was a between-subjects and within-story (i.e., within-item) factor. Speech content (speech with spatial content; speech without spatial content) was a within-subject and within-story factor. Spatial content type (literal; metaphorical with spatial language; metaphorical without spatial language) was a within-subject and between-story factor. When testing for interactions, we included random slopes only for the highest-order combination of within-unit factors subsumed by each interaction (Barr, 2013; e.g., by story random slope for Gesture Condition x Speech Content interaction). We used likelihood ratio tests (LRTs) to test for fixed effects, with post hoc contrasts performed on subsets of the data. When the models failed to converge, we simplified the random effects structure of the maximal model by (a) dropping the correlation between the random intercept and random interaction slope, (b) dropping the intercepts (Barr et al., 2013), (c) dropping the random interaction slope, and (d) dropping any random slope (i.e., intercept only model).

2.3 Results

2.3.1 Disfluency Rate

To compare speech disfluency rate across experimental conditions, we used mixed-effects Poisson regressions. We incorporated the number of words as an offset term into the model so that we modeled speech disfluency rate (number of speech disfluencies per word), rather than raw count data of number of speech disfluencies (Agresti, 2003). Overall, Poisson regression with an offset term allowed us to model rate data with successfully approximating a normal distribution and constant variance of residuals.

Overall Effect of Gesture Prevention on Disfluency Rate

Did people produce a higher rate of disfluencies when they were prevented from gesturing, compared with when they were allowed to gesture? In a first analysis including all speech content types (with and without spatial content), we found no evidence of an effect of gesture prevention on rate of speech disfluencies. Disfluency rates when people were prevented from gesturing ($M = .07, SD = .04, Median = .07$) were statistically indistinguishable from disfluency rates when people were allowed to gesture ($M = .06, SD = .03, Median = .06; \chi^2 = 1.07, p = .30$).

Effect of Gesture Prevention During Speech With Versus Without Spatial Content

A second analysis tested the effect of gesture prevention on disfluency rates in speech with spatial content and in speech with no spatial content. Results showed that preventing people from gesturing had no significant effect on disfluency rates during speech with spatial content ($\chi^2(1) = 1.71, p = .19$) or during speech with no spatial content ($\chi^2(1) = .56, p = .45$). Notably, the nonsignificant trends went in the opposite direction of what the LRH would predict: People were slightly less disfluent when prevented from gesturing compared with when they are allowed to gesture, both during speech with spatial content and during speech with no spatial content (see Figure 2.1). The (non-)effect of gesture prevention did not differ significantly between speech with and without spatial content, as indicated by a nonsignificant interaction between Gesture condition and Speech content ($\chi^2(1) = .30, p = .58$).

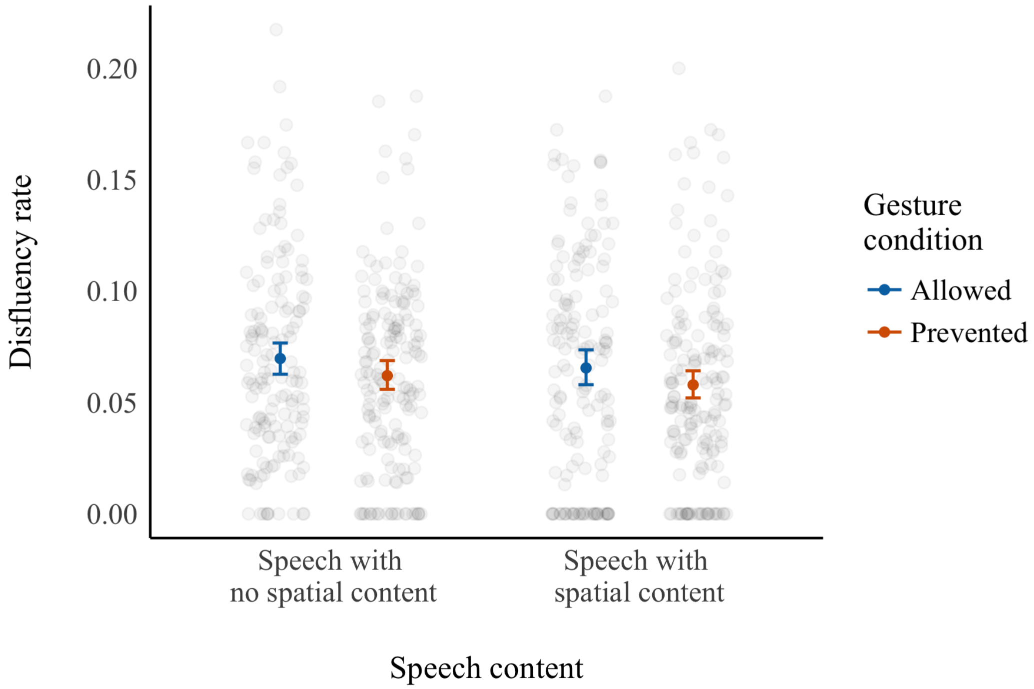

Figure 2.1: Disfluency rates during speech with spatial content (right) and speech with no spatial content (left). Gray dots show individual data points. Error bars show bootstrapped 95% confidence interval (CI) around the group means.

Effect of Gesture Prevention for Literal Versus Metaphorical Spatial Content

A third set of analyses tested for effects of preventing gesture during speech with literal spatial content and with metaphorical spatial content. Results showed no significant effect of gesture prevention on disfluency rate for any type of spatial content (Literal: $\chi^2(1) = .97, p = .32$; Metaphorical with spatial language: $\chi^2(1) = 1.56, p = .21$; Metaphorical without spatial language: $\chi^2(1) = 1.09, p = .30$), and the (non-)effect of gesture prevention on disfluency rate did not differ across these conditions ($\chi^2(1) = .07, p = .97$; see Figure 2.2).

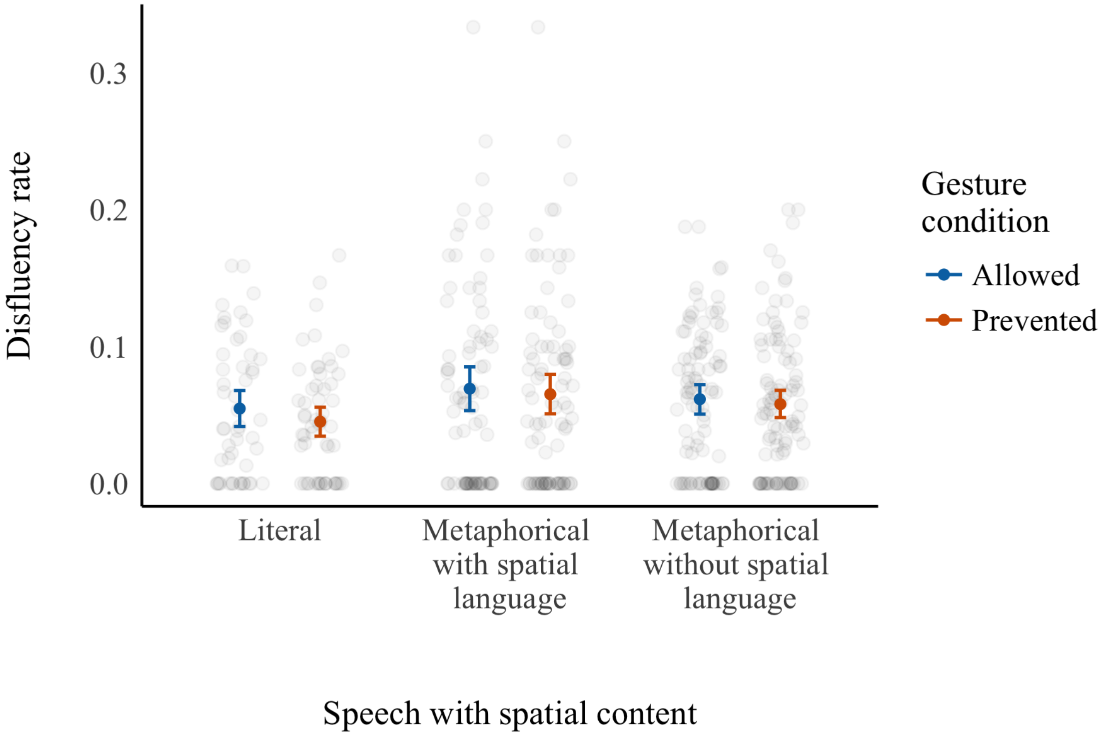

Figure 2.2: Disfluency rates during speech with literal and metaphorical spatial content. Gray dots show individual data points. Error bars show bootstrapped 95% confidence interval (CI) around the group means.

Effect of Gesture Prevention for Different Disfluency Types

Did people produce a higher rate of any kind of disfluency (i.e., repairs, repeats, filled pauses, or unfilled pauses) when they were prevented from gesturing, compared with when they were allowed to gesture? There was no effect of preventing gesture on the rate of any type of disfluency: repairs ($\chi^2(1) = .04, p = .85$), repeats ($\chi^2(1) = 1.01, p = .31$), unfilled pauses ($\chi^2(1) = .02, p = .89$), or filled pauses ($\chi^2(1) = 1.05, p = .30$). Similarly, there was no effect of preventing gesture on disfluency rates in speech with spatial content across different disfluency types: repairs ($\chi^2(1) = .54, p = .46$), repeats ($\chi^2(1) = .49, p = .48$, unfilled pauses ($\chi^2(1) = .11, p = .74$), or filled pauses ($\chi^2(1) = .85, p = .35$). Notably, the nonsignificant trends went in the opposite direction of what the LRH would predict: For any kind of disfluency, people were slightly less disfluent when prevented from gesturing compared with when they are allowed to gesture during speech with spatial content (see Figure 2.3).

2.3.2 Speech Rate

To compare speech rate for each phrase across experimental conditions, we used mixed-effects Gaussian regressions.

Overall Effect of Gesture Prevention on Speech Rate

Did people produce slower speech when they were prevented from gesturing, compared with when they were allowed to gesture? In a first analysis including all speech content types (with and without spatial content), we found no evidence of an effect of preventing gesture on speech rate. Speech rates when people were prevented from gesturing ($M = 212.98, SD = 86.91$) were statistically indistinguishable from speech rates when people were allowed to gesture ($M = 211.25, SD = 85.53, \chi^2(1) = .08, p = .77$).

Effect of Gesture Prevention During Phrases With Versus Without Spatial Content

We tested the effect of preventing gesture on speech rate in phrases with spatial content and in phrases with no spatial content. Results showed that preventing people from gesturing had no significant effect on people's speech rates during phrases with spatial content ($\chi^2(1) = .001, p = .97$) or during phrases with no spatial content ($\chi^2(1) = .23, p = .63$). The (non-)effect of gesture prevention did not differ significantly between phrases with and without spatial content, as indicated by a nonsignificant interaction between Gesture condition and Speech content ($\chi^2(1) = 1.36, p = .24$; see Figure 2.4).

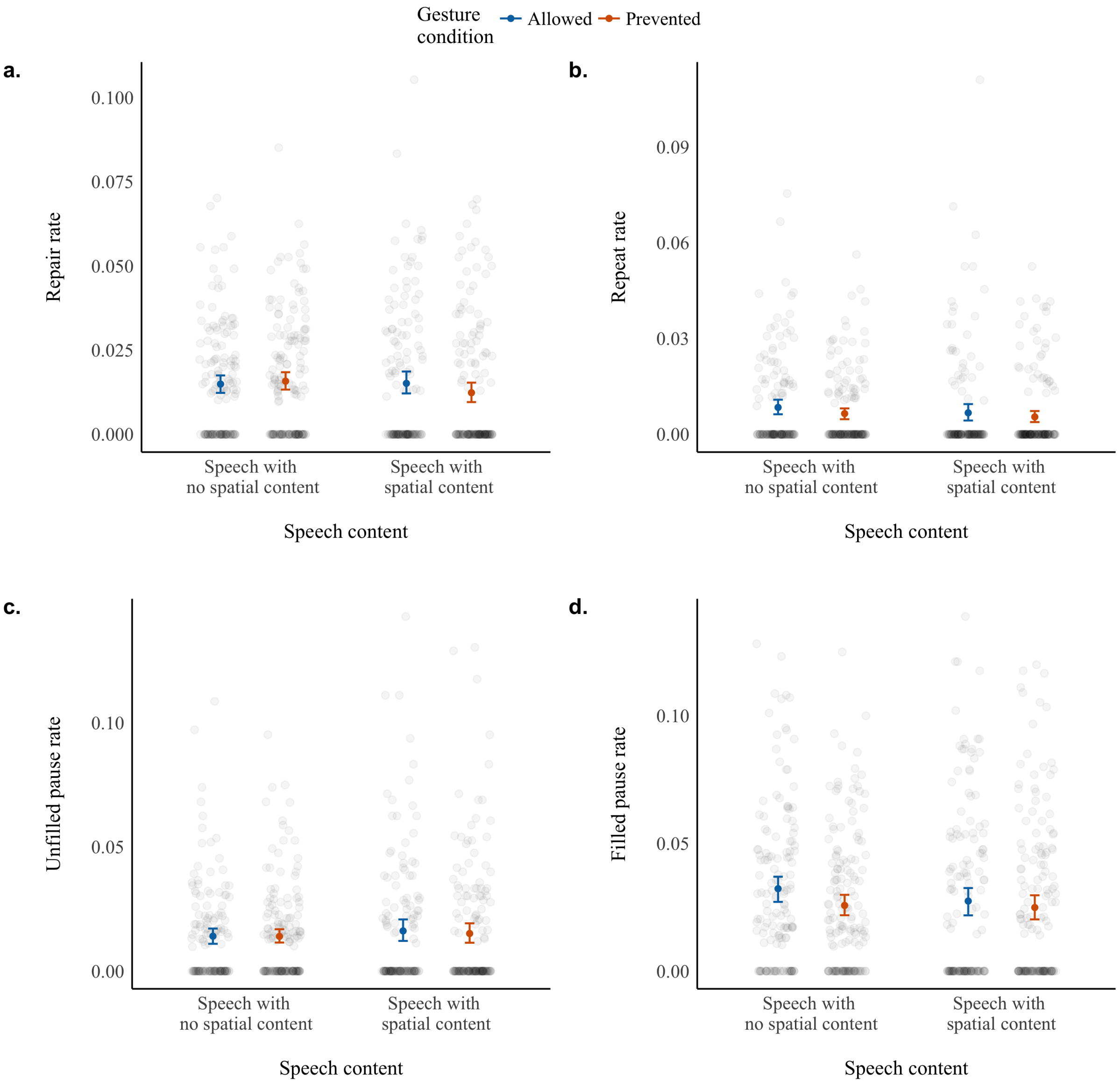

Figure 2.3: Disfluency rates during speech with spatial content and speech with no spatial content across different types of disfluency: (a) repairs, (b) repeats, (c) unfilled Pauses, and (d) filled Pauses. Gray dots show individual data points. Error bars show bootstrapped 95% confidence interval (CI) around the group means.

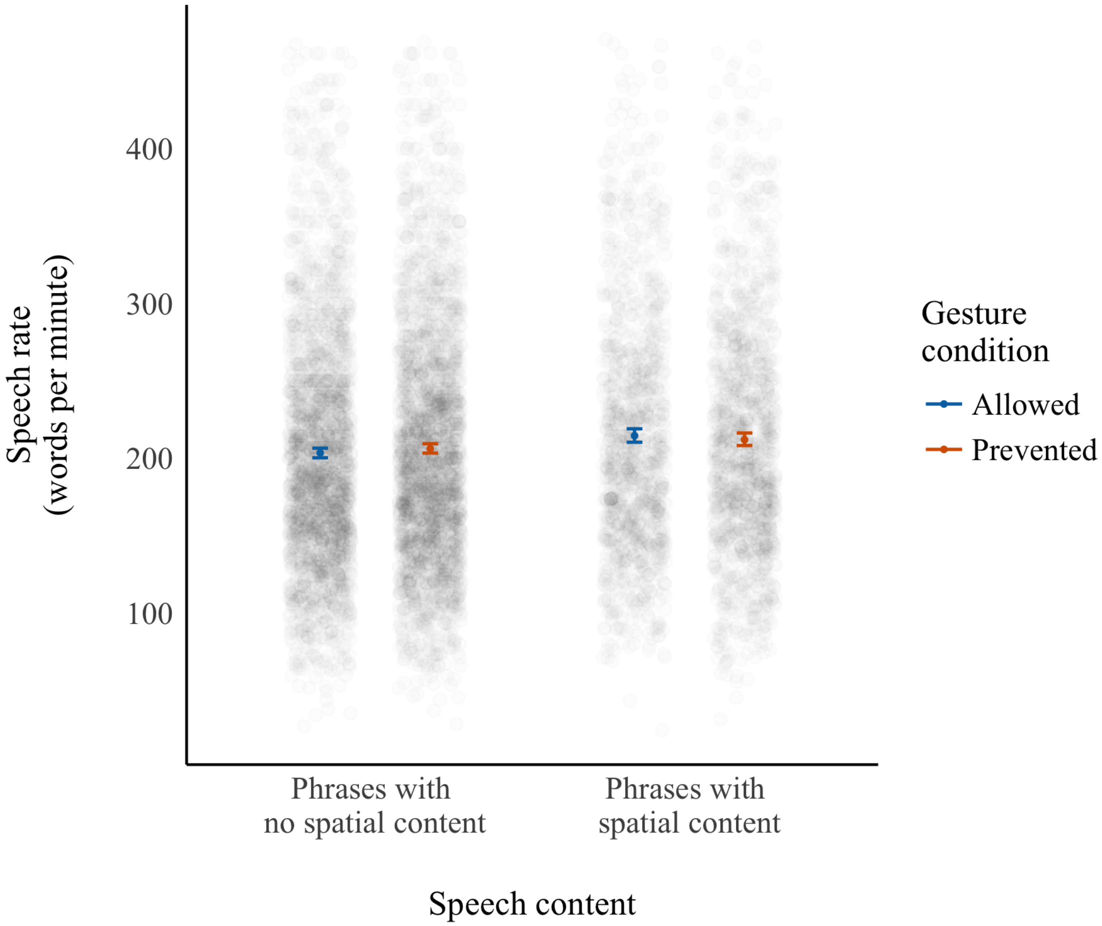

Figure 2.4: Speech rates (in words per minute) during phrases with spatial content (right) and phrases with no spatial content (left). Gray dots show individual data points. Outliers (data points that are 3 SDs above the mean) are removed for visualization purposes. Error bars show bootstrapped 95% confidence interval (CI) around the group means.

Effect of Gesture Prevention for Literal Versus Metaphorical Spatial Content

We tested for effects preventing gesture on speech rate during speech with literal spatial content and with metaphorical spatial content. Results showed no significant effect of preventing gesture on speech rate for any type of spatial content (Literal: $\chi^2(1) = .10, p = .75$; Metaphorical with spatial language: $\chi^2(1) = .42, p = .52$; Metaphorical without spatial language: $\chi^2(1) = .11, p = .74$), and the (non-)effect of gesture prevention on speech rate did not differ across these conditions ($\chi^2(1) = 1.16, p = .56$; see Figure 2.5).

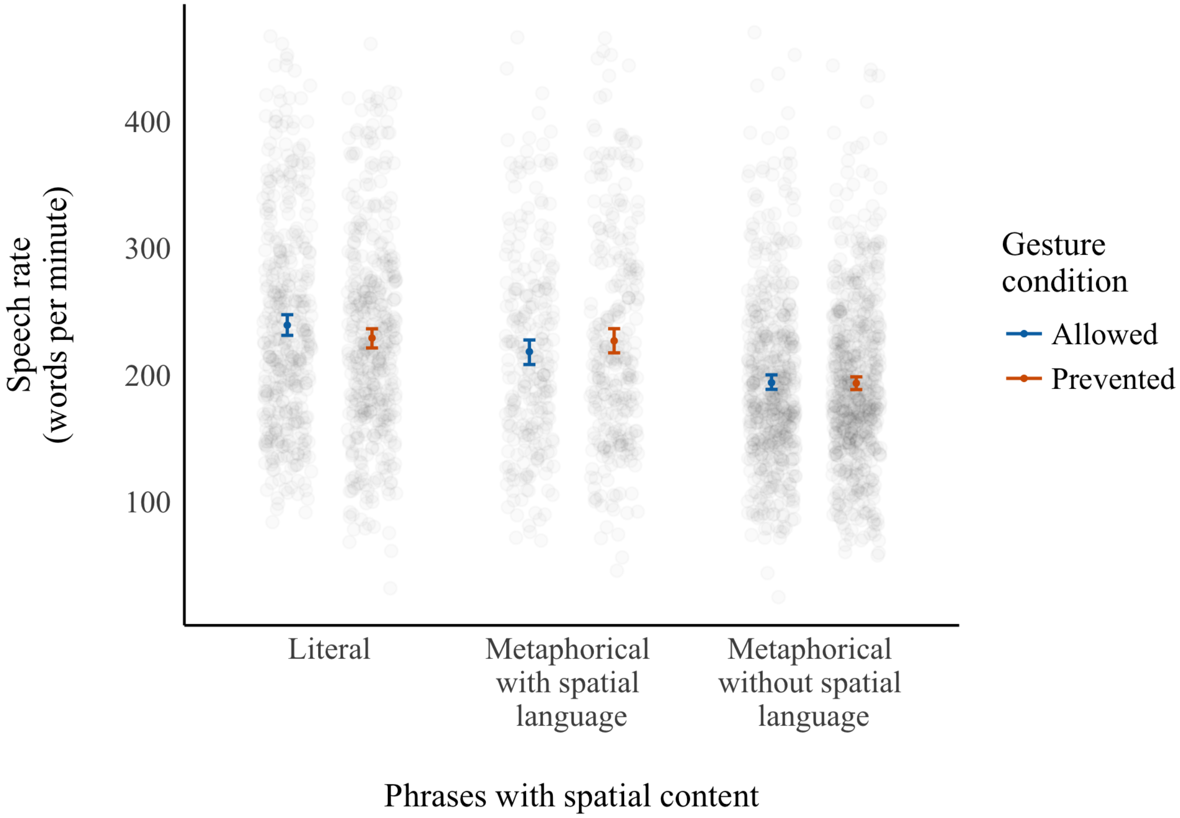

Figure 2.5: Speech rates (in words per minute) during phrases with literal and metaphorical spatial content. Gray dots show individual data points. Outliers (data points that are 3 SDs above the mean) are removed for visualization purposes. Error bars show bootstrapped 95% confidence interval (CI) around the group means.

2.3.3 Rate of Nonjuncture Filled Pauses

To compare nonjuncture filled pause rate across experimental conditions, we used mixed effects logistic regressions with nonjuncture or juncture as the binary outcomes for a filled pause.

Overall Effect of Gesture Prevention on the Rate of Nonjuncture Filled Pauses

Did people produce a higher rate of nonjuncture filled pauses when they were prevented from gesturing, compared with when they were allowed to gesture? In a first analysis including all speech content types (with and without spatial content), we found no evidence of an effect of preventing gesture on the rate of nonjuncture filled pauses. The nonjuncture filled pause rate when people were prevented from gesturing (184 nonjuncture filled pauses out of 434 filled pauses) was statistically indistinguishable from the same rate when people were allowed to gesture (216 nonjuncture filled pauses out of 459 filled pauses, $\chi^2(1) = .87, p = .35$; see Table 2.1).

Effect of Gesture Prevention During Phrases With Versus Without Spatial Content

We tested the effect of preventing gesture on nonjuncture filled pause rate in phrases with spatial content and in phrases with no spatial content. We found that preventing people from gesturing had no significant effect on people's nonjuncture filled pause rate during phrases with spatial content ($\chi^2(1) = .20, p = .65$) or during phrases with no spatial content ($\chi^2(1) = .47, p = .49$). The (non-)effect of gesture prevention did not differ significantly between phrases with and without spatial content, as indicated by a nonsignificant interaction between Gesture condition and Speech content ($\chi^2(1) = .29, p = .59$; see Figure 2.6)

Table 2.1

Number of non-juncture filled pauses (out of the total number of filled pauses) during the Gestures allowed and Gestures prevented conditions

	Gesture condition	
	Gestures allowed	Gestures prevented
Phrases with no spatial content	149 (287)	125 (266)
Phrases with spatial content:		
Literal	23 (48)	9 (33)
Metaphorical with spatial language	12 (34)	11 (29)
Metaphorical without spatial language	32 (90)	39 (106)
All phrases	216 (459)	184 (434)

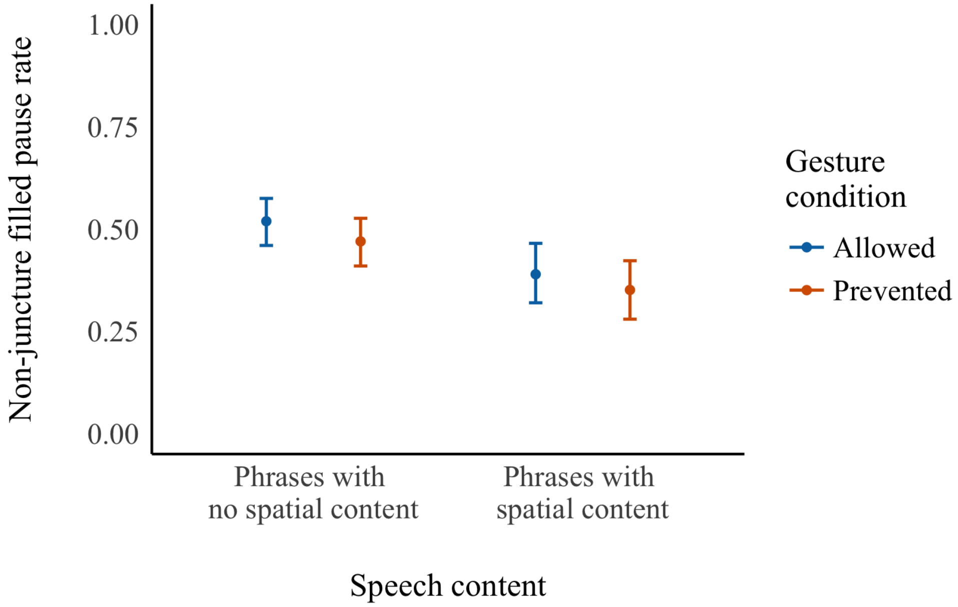

Figure 2.6: Nonjuncture filled pause rate (i.e., proportion of filled pauses that are nonjuncture as opposed to juncture) during phrases with spatial content (right) and phrases with no spatial content (left). Error bars show bootstrapped 95% confidence interval (CI) around the group means. For exact counts of filled pauses across conditions see Table 2.1.

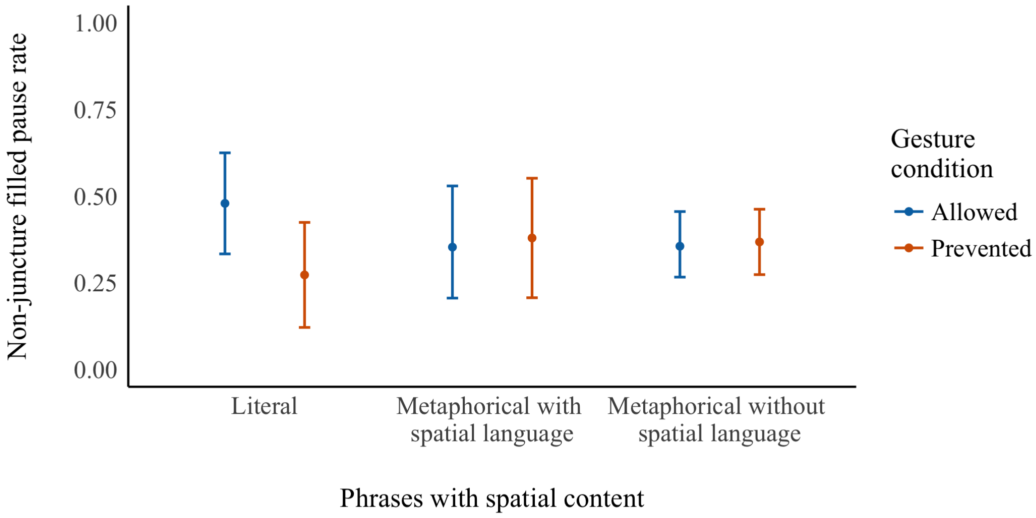

Figure 2.7: Nonjuncture filled pause rate (i.e., proportion of filled pauses that are nonjuncture as opposed to juncture) during phrases with spatial content. Error bars show bootstrapped 95% confidence interval (CI) around the group means. For exact counts of filled pauses across conditions see Table 2.1.

Effect of Gesture Prevention for Literal Versus Metaphorical Spatial Content

We tested for effects of preventing gesture on nonjuncture filled pause rate during speech with literal spatial content and with metaphorical spatial content. The effect of gesture prevention on nonjuncture filled pause rate was not significant for phrases with metaphorical spatial content (Metaphorical with spatial language: $\chi^2(1) = .06, p = .81$; Metaphorical without spatial language: $\chi^2(1) = .05, p = .82$). This effect was nominally significant (i.e., $p, .05$) for phrases with literal spatial content (Literal: $\chi^2(1) = 4.13, p = .04$); however, the effect went in the opposite direction from what the LRH would predict: People produced a lower rate of nonjuncture filled pauses when prevented from gesturing compared with when they are allowed to gesture. We do not interpret this backward result, for two reasons. First, this result is “significant” only if we maintain a nonconservative alpha value (.05) that is not corrected for multiple comparisons. Alpha correction would be necessary, especially for interpreting a single nonpredicted result, given that our search for any significant effect of gesture prevention led us to conduct 34 independent tests (22 planned, 12 post hoc); after appropriate correction, this nonpredicted result is no longer statistically significant. Second, this nonpredicted effect for phrases with literal spatial content did not differ significantly from the (non-)effects of gesture prevention for phrases with metaphorical spatial content, as indicated by a nonsignificant interaction between Gesture condition and Spatial content type ($\chi^2(1) = 2.92, p = .23$; see Figure 2.7).

2.4 Discussion

Does gesturing facilitate speech production by helping people find the right spatial words? We found no evidence that speakers speak less fluently when they are prevented from gesturing, compared with when they are allowed to gesture freely. Preventing gesture did not have the predicted effect on disfluency rate, speech rate, or nonjuncture filled pause rate overall, or on any of the experimental conditions we evaluated. We failed to find an effect of preventing gesture during speech with metaphorical spatial content, a finding that would have expanded the scope of LRH to encompass speech about abstract concepts. More fundamentally, we also found no significant effect of preventing gesture during speech with literal spatial content. We failed to find support for Rauscher and colleagues' (Rauscher et al., 1996) influential claim that preventing gesture increases disfluencies for spatial language. More broadly, our data provide no support for the idea that gesturing facilitates speech production by helping people find the right words.

2.4.1 Why Did We Find No Evidence That Gestures Help Speakers Find the Right Spatial Words?

Why did preventing gestures have no effect on our participants' speech, even when speakers were using words like “up” and “down” to describe concrete spatial scenarios? A first possible explanation to consider for any null result may be lack of statistical power. This explanation is unlikely for our study, however, given (a) the size of our data set, (b) our analysis choices that leverage the large number of within-subject observations, and (c) the qualitative patterns in the data. Overall, we analyzed 7,969 phrases containing 2,075 disfluencies, resulting in a large number of observations per subject (e.g., 156 observations per subject in the speech rate analysis). Unlike Rauscher and colleagues' (Rauscher et al., 1996), we did not average across observations for each subject, but instead conducted our statistical analyses for each story, clause, or phrase the participants produced. Modeling all data points increased our statistical power by increasing within-subject observations, while also reducing a source of Type I error (i.e., false positives) that was present in Rauscher et al.'s (1996) analyses by accounting for item-wise variance (Baayen et al., 2008; Clark, 1973a). Finally, if a lack of power were responsible for our null effects, then, overall, we would expect to find trends in the predicted direction that failed to reach statistical significance. This was not the case. On the contrary, as noted above, the strongest trends went opposite from the predicted direction, including the only trend among our 34 comparisons that was "significant" at $p < .05$ before correction for multiple comparisons. Considering all of the between-condition comparisons shown in Figures 2.1–2.7 (i.e., comparing each pair of adjacent light gray and dark gray bars), about twice as many comparisons trended against the LRH as trended in support of it.

A second possible explanation for a null effect could rest in having the predicted effects "hidden" in subsets of the data and obscured by aggregating over conditions or trial types. To ensure that this was not the case, we report graphs and planned analyses for all subsets of the data, broken out not only by spatial versus nonspatial content, but also by multiple types of language (literal, metaphorical spatial, and nonspatial language). To ensure that effects of any particular type of disfluency were not being masked by noneffects for other types, we also conducted post hoc analyses of each disfluency type, individually (Figure 2.3); preventing gestures did not have the predicted effect on any type of disfluency.

A third possible explanation for our null results could rest in levels of difficulty in speech production. Might effects of gesture prevention on fluency only emerge during difficult production conditions—and if so, is it possible that our production task was simply too easy? An examination of Rauscher et al.'s (1996) results does not support this possible explanation. Rauscher and colleagues (Rauscher et al., 1996) built into their experimental design three different levels of difficulty producing speech, positing that more difficult conditions should increase disfluency. Their participants produced natural speech with normal demands on word production in one condition (normal-speech condition); in a second condition participants had to use as many obscure words as possible (obscure-speech condition); in a third condition they had to avoid using words that contained a specified letter (constrained-speech condition). The latter two conditions make higher demands on word production than the normal-speech condition, and produced higher disfluency rates, overall.

Did preventing gestures increase disfluencies more when speech production was more difficult? Rauscher and colleagues (Rauscher et al., 1996) did not report any analyses that addressed this question, but the trends shown in their plots do not support the possibility that greater production difficulty increases the effect of gesture prevention on disfluency (see Figures 2.8 and 2.9 below, reprinted from Rauscher et al., 1996). On the contrary, for nonjuncture filled pauses (Figure 2.9) the effects of gesture prevention in Rauscher et al. (1996) were numerically smaller when the speech production task was harder. Speech production in our task was clearly less challenging than in Rauscher et al.'s (1996) obscure speech and constrained speech conditions; it is hard to know whether speech production in our task was more or less difficult than in Rauscher et al.'s "natural" speech condition, but Rauscher et al.'s results provide no evidence that making a speech production task more difficult yields a more sensitive test of effects of gesture prevention on disfluency.

Another possible explanation for our null results could rest in how complex the spatial ideas were that were expressed in speech. Might effects of gesture prevention on fluency

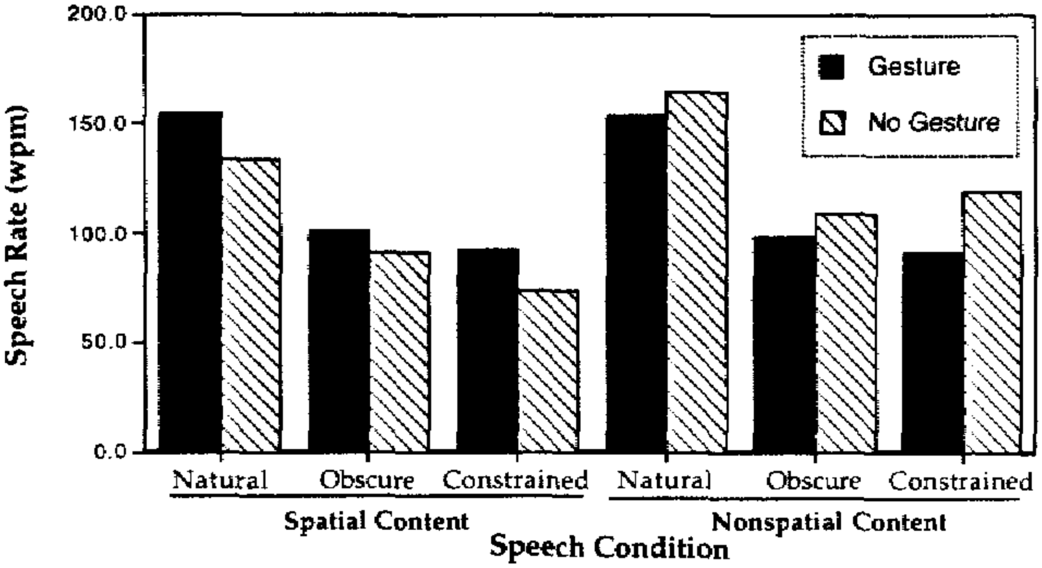

Figure 2.8: Speech rate (words per minute) in the natural-, obscure-, and constrained Speech conditions for speech with spatial and nonspatial content and when subjects were and were not allowed to gesture. Reprinted from "Gesture, Speech, and Lexical Access; The Role of Lexical Movements in Speech Production," by F. H. Rauscher, R. M. Krauss, and Y. Chen, 1996, Psychological Science, 7(4), p. 229. Copyright [1996] by SAGE Publications. Reprinted with permission.

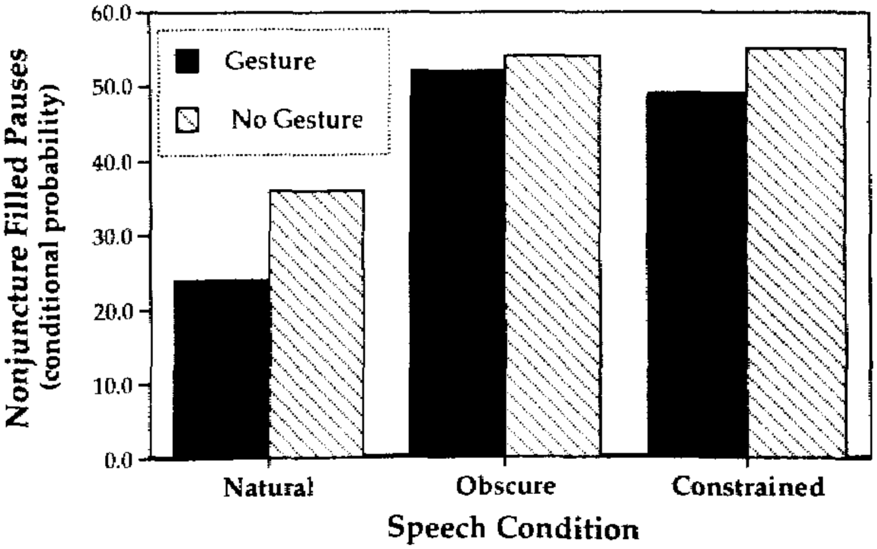

Figure 2.9: Probability of a nonjuncture filled pause given a filled pause in the natural-, obscure-, and constrained- speech conditions for speech with spatial and nonspatial content and when subjects were and were not allowed to gesture. Reprinted from "Gesture, Speech, and Lexical Access; The Role of Lexical Movements in Speech Production," by F. H. Rauscher, R. M. Krauss, and Y. Chen, 1996, Psychological Science, 7(4), p. 229. Copyright [1996] by SAGE Publications. Reprinted with permission.

only emerge during speech expressing complex spatial ideas—and if so, is it possible that our speech production task was semantically too simple? This explanation is unlikely to account for the null results in our study since the spatial descriptions in our stories were quite complex and varied involving position, shape, motion, and trajectory information (see online supplemental materials, e.g., story transcripts). The spatial words and phrases elicited in the current study cannot exhaust the range of complexity that could be present in a spatial description. Therefore, it is an open question whether gestures facilitate only highly complex spatial speech, beyond the level of complexity tested here.

The most fruitful explanation for the null results in the present study, we believe, rests in a reexamination of Rauscher et al.'s (1996) results, and of other studies testing the LRH. We first reexamine Rauscher et al.'s (1996) study in extensive detail to evaluate their proposed evidence for the LRH, because this is the only study in the gesture literature that claims to have found evidence that gesture prevention makes spatial speech disfluent, a claim that we failed to replicate here, and is one of the most cited studies in the gesture literature that largely shaped theories of why people gesture when they speak. In aiming to answer the question whether gestures facilitate speech production, it is critical to reexamine this seminal study by Rauscher et al. (1996) that has been the major source of answer to this question since its publication. We then examine in detail other studies that aimed to test effects of gesture prevention on speech disfluency. Much to our surprise, a careful examination of the highly influential study by Rauscher et al. (1996) and other studies that tested effects on speech disfluency yields no clear evidence that preventing gesture increases disfluencies in speech.

2.4.2 Is There Any Evidence That Gestures Help Speakers Find the Right Spatial Words?

Rauscher and colleagues' (Rauscher et al., 1996) study has been widely cited as evidence for the idea that gesturing helps speakers find the right words and, more broadly, as some of the first evidence that gesturing serves a cognitive function for speakers (e.g., Alibali et al., 2000, 2011; Casasanto, 2013; Goldin-Meadow, 1999; Hoetjes et al., 2014; Hostetter, 2011; Iverson & Goldin-Meadow, 1998). Did Rauscher et al. (1996) find the pattern of results predicted by their hypothesis—that is, did their study show that gesturing helps speakers find the right spatial words? No. To support their conclusions, Rauscher and colleagues would need to have shown particular patterns of data in each of their three dependent measures (i.e., speech disfluency rate, speech rate, and nonjuncture filled pause rate). Below, we outline the predicted patterns for each of these measures and explain why the observed patterns did not support these predictions.

To preview these explanations, in each of the three dependent measures the LRH predicted (and required) two effects: First, preventing gesture should make spatial speech more disfluent, resulting in a simple effect of gesture prevention in speech with spatial content. Second, preventing gesture should increase disfluency selectively during speech with spatial content, as opposed to speech with no spatial content, resulting in a two-way interaction of gesture condition (gestures allowed; gestures prevented) and speech content (spatial content; no spatial content). The selectivity of this interaction effect is crucial for the data to support the LRH, which hypothesizes that there is a special link between gesture and spatial words. Furthermore, without this two-way interaction, the effect of gesture prevention would be open to a hypothesis-irrelevant interpretation: Preventing gesture on any kind of speech could increase disfluency simply due to the unnaturalness of preventing gesture (see Rauscher et al., 1996, p. 229 for a similar argument).

Speech Disfluency Rate

Rauscher et al.'s (1996) most influential claim is that preventing gesture causes higher disfluency rates only during speech with spatial content. Yet, the simple effect required to support this claim was not statistically significant, and the required twoway interaction was never reported, neither in Rauscher et al. (1996), nor in subsequent review articles and chapters highlighting these results (see Figure 4 in Krauss, 1998; reprinted here as Figure 2.10). Turning first to the required simple effect, the LRH predicted that participants should produce more disfluencies when they are prevented from gesturing, compared with when they are allowed to gesture, during speech with spatial content (i.e., the two bars on the left in Figure 2.10 should be significantly different from each other). However, this critical simple effect was only marginally significant, as reported (i.e., $p < .066$). Notably, even this reported value is anticonservative, in at least two ways. First, the statistical test did not account for item-wise variance (Baayen et al., 2008; Clark, 1973a), leading to an increased probability of Type I error (i.e., a false positive result). Second, the alpha value for pairwise comparisons was not corrected for the multiple statistical comparisons reported, and the even greater number of comparisons that could have been conducted in this 2 x 2 x 3 design. After correcting the alpha value appropriately, the reported marginal p-value would no longer approach significance.

This simple effect was necessary to support Rauscher et al.'s (1996) main claim; because it was not significant, there is no need to analyze the speech disfluency rates results further to evaluate their significance. However, as noted, there was a second effect required by the LRH as well: the two-way interaction of gesture condition and speech content. The significance of this critical interaction was implied by Rauscher et al.'s (1996) text, but a different interaction was reported in its place: the three-way interaction of gesture condition (gesture allowed; gesture prevented), speech content (spatial content; nonspatial content), and speech condition (natural = producing natural speech; obscure = using as many obscure words as

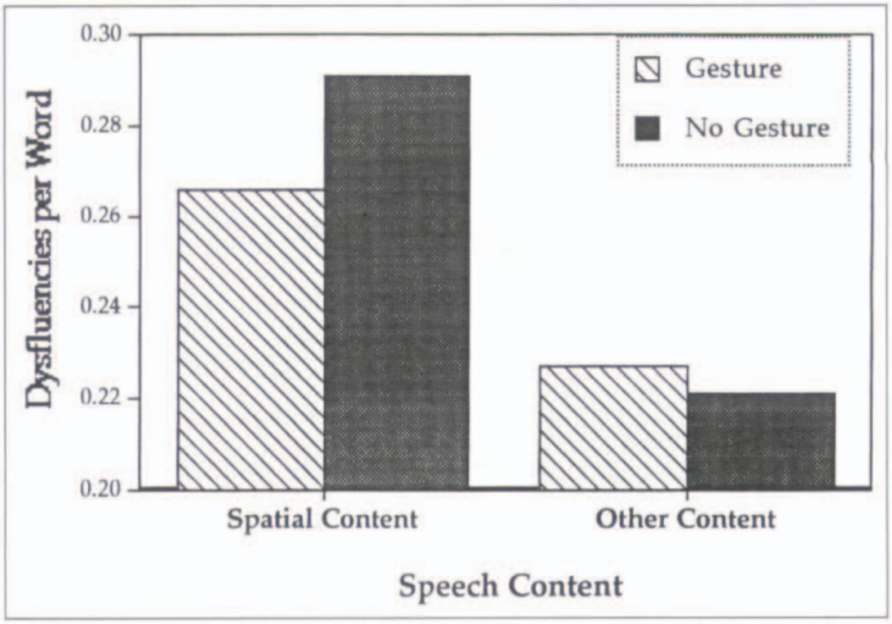

Figure 2.10: Disfluency rates for speech with spatial content and with nonspatial content and when subjects were and were not allowed to gesture. Reprinted from “Why do we gesture when we speak?”, by R. M. Krauss, 1998. Current Directions in Psychological Science, 7(2), p. 58. Copyright [1998] by SAGE Publications. Reprinted with permission.

possible; constrained = avoiding using words that contained a specified letter). Although this three-way interaction was significant, it is irrelevant to testing the LRH, and it does not support the author’s claim that “the effects of preventing gesturing depended on whether the conceptual content of the speech was spatial or nonspatial” (p. 229).

Is it possible that Rauscher and colleagues (Rauscher et al., 1996) found the selective effect of gesture prevention on spatial speech, but simply failed to report the required two-way interaction? This was not the case; given the nonsignificance of the critical simple effect of preventing gesture in the spatial speech condition, the only way that the two-way interaction could become significant would be if preventing gesture had an unpredicted facilitating effect on speech production in the nonspatial speech condition. Thus, even if the nonreported (but critical) two-way interaction were significant, this interaction would not provide any clear support for the LRH since (a) it did not comprise the critical simple effect in the spatial speech condition, and (b) the statistical significance of the interaction would depend on a "backward" simple effect in the nonspatial speech condition, where no effect of gesture prevention was predicted (see the two bars on the right in Figure 2.10).

Speech Rate

Rauscher et al. (1996) reported analyses of speech rate as a second test of the LRH. As for disfluency rate, however, there was no clear evidence that the two required effects of gesture prevention on speech rate supported the LRH. For speech rate, the LRH predicted a critical simple effect showing that preventing gesture causes slower *spatial* speech. Looking at the results depicted in their plots, there is a trend consistent with this simple effect: Numerically, the speech rate was slower during speech with spatial content when participants were prevented from gesturing (striped bars), compared with when they were allowed to gesture (black bars, see the three pairs of bars on the left of Figure 2.8, reprinted from Rauscher et al., 1996). However, this trend may not be statistically significant; although the authors suggested that "with spatial content, speakers spoke more slowly when they could not gesture" (p. 228), no statistical test of this effect was reported, and no error bars were provided to guide interpretation of the trends.

Even if Rauscher et al. (1996) had obtained the critical simple effect, a further two-way interaction would be necessary to show that preventing gesture causes slower speech rates *selectively* during speech with spatial content. Rauscher et al. (1996) indeed reported a significant two-way interaction between gesture condition and speech content for speech rate. However, this interaction does not provide any clear support for the LRH since its composition appears problematic, in two ways: First, as noted above, there is no evidence that the required simple effect in the spatial speech condition was significant; second, the statistical significance of the interaction is driven in part by a trend toward a backward simple effect in the nonspatial speech condition, showing that people spoke *faster* when they were prevented from gesturing (i.e., see the three pairs of bars on the right of Figure 2.8 where the striped “no gesture” bars in Figure 2.8 are higher than the black “gesture” bars for speech with nonspatial content). Rauscher et al. (1996) acknowledged that a facilitating effect of preventing gesture on nonspatial speech was not predicted, writing: “when the content was nonspatial, speakers spoke more rapidly when they could not gesture. This latter result is puzzling to us, and we have no explanation for it” (p. 228).

In summary, the speech rate data from Rauscher et al. (1996) provide no clear support for the LRH. No statistical test was reported for the critical simple effect of gesture prevention on spatial speech. Although the critical two-way interaction was reported to be significant, (a) this interaction is necessary but not sufficient to support the LRH (absent the required simple effect in the spatial speech condition), and (b) its statistical significance is driven, in part, by the backward effect in the nonspatial speech condition.

Nonjuncture Filled Pause Rate

The third dependent measure reported by Rauscher et al. (1996) was the rate of nonjuncture filled pauses, which the authors suggest should be “the measure that most sensitively reflects problems in lexical retrieval” (p. 229) and is the most sensitive test of the LRH. On the basis of their analyses, Rauscher et al. (1996) claimed that “preventing gesturing increased the relative frequency of nonjuncture filled pauses in speech with spatial content, but not in speech with other content” (Rauscher et al., 1996, p. 226). This claim would require the same simple and interaction effects to be significant as in the previous dependent measures. In (partial) support of the LRH, the required simple effect was reported to be significant: Participants produced nonjuncture filled pauses at a higher rate when they were prevented from gesturing, compared with when they were allowed to gesture, during speech with spatial content. Yet, this simple effect is not sufficient to support the LRH. To support the claim that preventing gesture increases nonjuncture filled pauses *selectively*, “in speech with spatial content, but not in speech with other content,” it would be necessary to present data in both the spatial and nonspatial speech conditions, and to test for the required 2-way interaction of gesture condition (gesture allowed; gesture prevented) and speech content (spatial content; nonspatial content). This interaction was not reported, and it cannot be tested in the reported data because Rauscher et al. (1996) did not report any results concerning nonjuncture filled pause rates in the nonspatial speech condition (Figure 2.9, reprinted from Rauscher et al., displays only spatial speech). As such, the authors did not present or analyze the required pattern of results needed to support their conclusion about the selective effect of preventing gesture on spatial speech.

Beyond Rauscher and Colleagues

Several studies in the 5 decades before and after Rauscher et al. (1996) have also failed to find reliable evidence for an increase in disfluent speech when speakers are prevented from gesturing. To our knowledge, no study has found higher disfluency rates, or higher nonjuncture filled pause rates, when speakers are prevented from gesturing than when they are allowed to gesture (Cravotta et al., 2018; Finlayson et al., 2003; Graham & Heywood, 1975; Hoetjes et al., 2014; Hostetter et al., 2007; Rimé et al., 1984). Only two studies (out of seven) report overall slower speech rates when individuals are prevented from gesturing than when they are allowed to gesture (Cravotta et al., 2018; Morsella & Krauss, 2004). However, in one of these studies (Cravotta et al., 2018), the finding that speech was slower when gesture was prevented was the only significant comparison among a total of 22 comparisons, which included other disfluency measures; therefore, this single significant finding would not remain statistically significant after correcting the alpha-level for multiple comparisons. Morsella and Krauss (2004) also found slower speech rates when gesture was prevented than when it was allowed. However, the authors did not find a selective effect of gesture prevention on speech rate in one experimental condition more than in another; thus, it is unclear whether the results simply reflect a particularly restrictive gesture prevention method (i.e., participants had electrodes placed on their forearms and were instructed not to move because "movement of the limbs could ruin the quality of the recordings"; p. 89). Notably, none of the studies reporting null effects of gesture prevention distinguished between disfluencies during spatial and nonspatial speech; thus, arguably, these studies did not attempt to validate Rauscher et al.'s (1996) claim that gesture prevention selectively affects spatial speech. By contrast, our study tested this claim explicitly, but still found no evidence that gesture prevention increases disfluency—in either spatial or nonspatial speech.

2.4.3 Beyond Disfluency: Is There Any Evidence That Preventing Gestures Hurts Speech Production?

Given that there is no reliable evidence that gestures help people produce fluent speech, is there any evidence that gestures facilitate speech production in other ways? In principle, preventing gesture could hurt some aspect of speech production that does not result in speech disfluencies. For example, according to Rimé et al. (1984), preventing people from gesturing lowered the vividness of the imagery in speech. However, a reexamination of the analyses reported in Rimé et al. (1984) indicates that this reported effect should not be interpreted as statistically significant production. The results showed a p-value less than .05 for only one these six variables (effect of gesture prevention on imagery index: $F(4, 48) = 3.19, p = .02$). This result is widely cited as evidence that gestures facilitate speech production by activating semantic features that enter into word search (e.g., see Rauscher et al., 1996, p. 226). Yet, interpreting this one result in the context of Rimé et al.'s full study would require correcting the alpha level for multiple comparisons. A Bonferroni-corrected alpha-value for six independent tests would be $\alpha = .008$; Rimé et al.'s (1984) reported effect of gesture prevention on imagery would not approach significance based on this corrected alpha-value.

Another set of studies testing effects of gesture prevention on speech, beyond disfluencies, examined whether people find it harder to generate a target word from definitions or from pictures when prevented from gesturing. This effect was tested in three papers (Beattie & Coughlan, 1999; Frick-Horbury & Guttentag, 1998; Pine et al., 2007). In influential papers reviewing these findings (e.g., see Cook et al., 2010; Hostetter & Alibali, 2008; Kita et al., 2017; Wesp et al., 2001), two out of the three studies have been cited as evidence that gestures help people find the right words; only one study, by Beattie and Coughlan (1999), has been cited as showing no evidence in support of this view. Our reexamination of these three studies revealed that, in fact, none of these papers shows consistent support for the hypothesis that gesturing helps people to find target words. Below, we examine the pattern of results that led Beattie and Coughlan (1999) to conclude that they found no evidence in support of the LRH, and we show that according to the same criteria, the two other studies showed no clear evidence in support of the LRH (Frick-Horbury & Guttentag, 1998; Pine et al., 2007). On the contrary, Frick-Horbury and Guttentag's (1998) study offered the opposite conclusion: "In the present study, however, there was little evidence that gesture production per se enhanced verbal recall" (p. 54).

What were the patterns of results that led Beattie and Coughlan (1999) to conclude that gestures did not help people find target words, as had been predicted? Beattie and Coughlan (1999) found no statistically significant evidence that people remember fewer words when prevented from gesturing, compared with when they were allowed to gesture—this null effect in Beattie and Coughlan (1999) has been acknowledged by subsequent papers summarizing these results. However, this null effect of gesture prevention on word finding was only one among a broader set of results that led Beattie and Coughlan (1999) to conclude that their study showed no evidence in support of the LRH, and it was not even the effect that the authors considered to be the most critical (see p. 41 and 50). They also tested whether gesture prevention makes it harder for people to resolve tip-of-the-tongue states (TOT states): situations in which speakers know the gist of what they want to say, but they cannot find the right word. Beattie and Coughlan (1999) did find statistically significant evidence that gesture prevention makes it harder to resolve TOT states, but they did not interpret this result as supporting the LRH. Critically, they argued that an effect of gesture prevention on resolving TOT states is not clear evidence for the LRH. Why not? Because only iconic gestures are hypothesized to "have a functional role in lexical access" (Beattie & Coughlan, 1999, p. 35; see also Butterworth & Hadar, 1989). If an analysis includes all gesture types, then the effects of gesture prevention on resolving TOT states "may, of course, have nothing to do with the occurrence of iconic gestures" (p. 41, italics added). In fact, there was no evidence to suggest that Beattie and Coughlan's (1999) effect was driven by the prevention of iconic gestures; therefore, the effect did not support the LRH (or related proposals, e.g., Butterworth & Hadar, 1989).

Beyond the ambiguity of their gesture prevention results, Beattie and Coughlan (1999) tested for two critical links between gesture production and word finding, both of which failed to support the LRH. First, if gestures help people resolve TOT states, then people should resolve more TOT states on trials where they produced a gesture, compared with trials where they did not produce a gesture. However, Beattie and Coughlan (1999) found the opposite pattern: Only 64% of the TOT states were resolved when people produced a gesture, whereas all of the TOT states were resolved when people did not produce a gesture. Second, the LRH predicts that only iconic gestures should help people resolve TOT states. If this is true, then more TOT states should be resolved when people produced iconic gestures, compared with when they did not produce any iconic gestures. Yet, there was no statistically significant evidence that people resolved more TOT states when they produced an iconic gesture (69%), compared with when they did not (73%); rather, there was a numerical trend in the opposite direction. Overall, even though Beattie and Coughlan (1999) obtained one result predicted by the LRH, a main effect of gesture prevention on TOT resolution, they concluded that "if one considers the broader context here, there is no real evidence that gesturing facilities lexical access" (p. 47).

Although Frick-Horbury and Guttentag (1998) is widely cited as evidence for the LRH, the authors stated that their study does not provide clear support for the LRH because they found that the occurrence of gestures was not actually associated with successful word finding. Frick-Horbury and Guttentag (1998) did find a main effect of gesture prevention: People remembered fewer words when they were prevented from gesturing, compared with when they were allowed to gesture. However, like Beattie and Coughlan (1999), Frick-Horbury and Guttentag (1998) argued that this effect of gesture prevention on word finding is not clear evidence for the LRH. To determine whether their results supported the LRH, they tested whether gesture production was associated with successful word finding. Like Beattie and Coughlan (1999), Frick-Horbury and Guttentag (1998) found the opposite pattern: Only 21% of the words were correctly remembered when people produced a gesture, whereas 53% of the words were correctly remembered when people did not produce a gesture. Based on these results, FrickHorbury and Guttentag (1998) concluded that "It is possible that a focus on overt gesture production is the wrong place to look for an explanation for the hand restriction effect."

Using the same criteria that Beattie and Coughlan (1999) and FrickHorbury and Guttentag (1998) used in evaluating their results, we can conclude that Pine et al. (2007) also did not provide clear evidence in support of the LRH. Pine et al. (2007) did find two main effects of gesture prevention: Children remembered fewer words, and resolved fewer TOT states, when they were prevented from gesturing, compared with when they were allowed to gesture. And unlike the other studies, Pine et al. (2007) did also find that gesture production may be responsible for these effects: Children gestured more before successfully finding a target word and before successfully resolving a TOT state. However, critically, they showed no evidence to suggest that these effects were selectively driven by the production of iconic gestures: Children used more iconic gestures, but also more beat gestures and more self-adaptors (i.e., simple self-touching movements), when they were able to find a word (i.e., remembering the target word, resolving a TOT state), compared with when they failed to find a word (i.e., not remembering the target word, failing to resolve a TOT state). Given that the association of more gestures with successful word finding was not selective to iconic gestures, iconic gesture production may be the wrong place to look for an explanation of the gesture prevention effects. Being allowed to make any movement, including self-adaptors, may make children feel more comfortable, which could help them succeed in the experimental task. Therefore, Pine et al.'s (2007) results also do not provide clear evidence for the LRH.

So far, we have focused our reexamination of Beattie and Coughlan (1999), Frick-Horbury and Guttentag (1998), and Pine et al. (2007) on only two measures of word finding: the number of target words remembered, and the number of TOT states resolved. However, all three studies also tested the effect of gesture prevention on a third measure of word finding: whether people experience more TOT states when they are prevented from gesturing, compared with when they are allowed to gesture. If gestures help people find the right words, people should experience more TOT states when gestures are not available to them. However, none of the three studies showed evidence supporting this prediction. Furthermore, Beattie and Coughlan (1999) found statistically significant evidence in the opposite direction of this prediction: People experienced fewer TOT states when they were prevented from gesturing, compared with when they are allowed to gesture. Overall, our reexamination revealed that none of these studies showed clear support for the LRH; on the contrary, some measures showed consistent evidence against the LRH.

To conclude, our reexamination of the five decades of research cited as evidence for the LRH corroborates our reexamination of Rauscher et al. (1996): There appears to be no clear evidence that preventing gesture hurts speech production—by making speech more disfluent, by lowering the imagery level in speech, by hindering word finding and TOT state resolution, or by making people experience more TOT states. Overall, we conclude that there is no clear support for the long-standing, influential claim that people gesture when they speak, in part, because gesturing helps speakers produce the right words.

2.4.4 Why Do People Gesture When They Speak?

Krauss (1998) offered the LRH as an answer to the question, “Why do we gesture when we speak?” A “why” question explores the function of a behavior: What consequences follow a behavior that leads to that behavior’s recurrence? The LRH provided the first functional explanation for what causes a speaker to gesture that did not restrict cospeech gesture’s function to communicative benefits for the listener, and instead suggested that gesturing may serve a cognitive function in the speaker’s mind.

Even though gesturing does not facilitate speech production, gestures do appear to serve a variety of other speaker-internal cognitive functions. For example, gesturing lightens speakers’ cognitive load and frees up cognitive resources that can be allocated to other tasks: Gesturing while explaining a math task allowed speakers to remember more words on a simultaneously performed word recall task (Goldin-Meadow et al., 2001; Ping & Goldin-Meadow, 2010; Wagner et al., 2004). Gesturing also facilitates learning: Gesturing while learning a new math concept helps learners generalize (Goldin-Meadow et al., 2009; Novack et al., 2014; Wakefield et al., 2018) and retain (Cook et al., 2008) the knowledge they gained during instruction.

As Krauss (1998) argued, gesture’s function cannot be limited to communicative purposes. A complete functional account of why people gesture needs to include not only communicative functions for the listener, but also speaker-internal cognitive functions. In light of the evidence we present here, such an account of the cognitive functions of gesture will need to look beyond the role of gesture in speech production.

2.5 Conclusions

Do gestures facilitate speech production? Typically, reviews listing the cognitive functions of gesture start this list with the assertion that gesture helps people find the right words (e.g., Alibali, 2005; Goldin-Meadow, 1999; Kita et al., 2017). Here, however, we showed that there is no compelling evidence to support this influential hypothesis. Rauscher et al.'s (1996) claim that gesture prevention makes spatial speech disfluent is among the most widely cited empirical results in the gesture literature; yet, upon reexamining this study, we found that the data do not support the LRH. Our further reexamination of 5 decades of research testing versions of this hypothesis, before and after Rauscher et al.'s (1996) influential study, revealed that there is no reliable evidence that gesture prevention hurts speech production. Accordingly, the results from our study showed no statistically significant effects of gesture prevention on speech disfluency, for speech about literal or metaphorical space. Gestures do not appear to facilitate speech production, challenging long-held beliefs about why people gesture when they speak.

CHAPTER 3

A PRAGMATIC ACCOUNT OF WHY SPEAKERS GESTURE DURING DISFLUENT SPEECH

3.1 Introduction

People gesture when they speak, especially when speaking is difficult. People are more likely to gesture when their speech is disfluent, compared to when their speech is fluent (Akhavan et al., 2016; Butterworth & Beattie, 1978; Ragsdale & Silvia, 1982). Why? It is clear that some gestures are intended to be communicative, produced for the listener (Bavelas et al., 1992). However, gestures that occur when speaking is difficult have been widely interpreted as evidence that they must be for the speaker, helping their own cognitive functions, rather than for the listener. As noted in Chapter 2, Krauss' influential account argues that people gesture when they are disfluent because those gestures resolve speech difficulties and thus facilitate speech production (Krauss & Hadar, 1999). This account predicts that when people are prevented from gesturing, they should experience more speech difficulties. However, in conflict with this prediction, there seems to be no reliable evidence that preventing gestures impairs speaking (see Study 1; also Kısa et al., 2021).

If gestures do not help resolve speech difficulties, why then are people more likely to gesture when they are disfluent? Here, we propose an alternative explanation: People gesture when their speech is disfluent because gestures serve as a pragmatic signal to the listener by commenting on problems with presenting an utterance.

According to Herbert Clark's (1996) theory of communication, speakers signal through two tracks simultaneously. In the primary track, they refer to the official business – the topic being talked about. In the collateral track, speakers comment on the performance of speaking itself. One situation when speakers should comment on the act of speaking is when they encounter problems with speech production – that is, delays or mistakes in doing the official business of presenting an utterance. When speakers need extra time to plan their utterance or when they say the wrong word or phrase, they should give an account of the problem. There are many reasons why speakers might acknowledge and or explain that they are deviating from doing the official business of presenting an utterance. Commenting on speech problems can give the listener information about the speaker's production plan and ensure successful coordination in conversation timing (Holler & Levinson, 2019). Commenting on speech problems is also in line with the social expectations in conversing – conversation is a joint activity and speakers would be violating the principle of being cooperative if they do not acknowledge their deviation from the role they commit to as a speaker — presenting an utterance (Clark, 1996). Accordingly, speakers use speech disfluencies as signals to account for interruptions in their speech (e.g. repairs, filled pauses like "um").

Gestures could also be a pragmatic signal commenting on the speaker's difficulty presenting an utterance. People may be more likely to gesture when they are disfluent, compared to when they are fluent, because gestures can comment on problems with presenting an utterance in many ways: they can foreshadow an incoming interruption, signal the intent to continue speaking during an interruption, acknowledge an ongoing interruption, or signal their commitment to a fluent re-start.

The goal of this chapter is to first conceptually replicate the finding that people are more likely to gesture when they are disfluent than when they are fluent. We then test whether people gesture when their speech is less fluent because gestures serve as a pragmatic signal to the listener. If so, when the listener cannot see the speaker, the speaker's pragmatic motivation to gesture during disfluencies should disappear. People should be more likely to gesture when they are disfluent than when they are fluent only (or primarily) when the listener can see their gestures (e.g., face to face conversation). On the other hand, if gestures do not serve as pragmatic signals, but they are merely symptoms of speech difficulties, then the visibility of gestures should not matter: People should be likely to gesture more during disfluent speech regardless of whether their gestures are visible or not. To test this possibility, we recorded gestures and speech disfluencies (within-phrase pauses, repeats or repairs of words or phrases, and filled pauses) when people's gestures are visible and not visible to their listener.

3.2 Methods

3.2.1 Participants

The story retelling sessions from the 28 participants in the Gestures allowed condition of Study 1 (from Chapter 2) were also used for this study – as the Gestures visible condition in the current study. Additionally, 28 Stanford University undergraduates were recruited in pairs, and participated for course credit after giving informed consent – for the Gestures not visible condition in the current study.

3.2.2 Materials

The 12 stories in the current study were identical to the ones used in Study 1, with the only exception that 3 of the stories were replaced with new ones, only for the gestures not visible condition. This resulted in a total of 15 brief stories, each 50-100 words, implying motion or extension in one of four spatial directions: upward, downward, right, or left. Just like in Study 1, some of the stories had literal spatial content (e.g., "the rocket went higher"; "the scuba diver went down") and other stories had metaphorical spatial content (e.g., "my grades went higher"; "the price went down").

3.2.3 Procedure

The story-telling procedure was identical to Study 1: Participants took turns studying written stories and then retelling the stories to their partners. Just like in Study 1, after starting with a warm-up story, each participant re-told 6 stories in randomized order: 2 stories with literal spatial content and 4 stories with metaphorical spatial content.

The group of participants who were assigned to the gestures allowed condition in Study 1 constituted the gestures visible condition in the current study, where participants were seated facing one another across a table. An additional group of participants were assigned to the gestures not visible condition: The listener was blindfolded and the participants were separated by an opaque barrier on the table's surface occluding gesture space. Testing lasted 20-30 minutes.

3.2.4 Coding

Clause coding

Participants' audio recordings of the stories in the Gestures visible condition were already transcribed verbatim and parsed into clauses for Study 1. For the current study, we did the same for the stories in the Gestures not visible condition: Coder 1 transcribed and parsed the stories into clauses. Participants retold a total of 336 stories across the two visibility conditions. The video recording for 3 of the stories are missing, so for the further analyses, we worked with 333 stories in total. Participants produced a total of 3534 spoken clauses (Gestures visible: 1936; Gestures not visible: 1598).

Each clause was classified as a spatial clause if it contained language that implied literal or metaphorical motion, length, or position along either the lateral or vertical axis. For instance, "my rocket went higher" would be a spatial clause. For each spatial clause, the spatial direction was coded to be upward, downward, leftward or rightward.

A total of 360 spoken clauses were excluded from further analyses: (i) Gestures were not codable for 108 of these clauses due to hands being occluded from view and (ii) disfluencies were not codable for 252 of these clauses due to low audio quality of the video recording. As a result, a total of 3174 clauses were included in the further coding and analyses.

Speech disfluency coding

Speech disfluencies were already coded for the Gestures visible condition in Study 1. Coder 2 (same disfluency coder with Study 1) recorded the location and type of speech disfluency for each clause, for the Gestures not visible condition – using only the audio with no video. Speech disfluencies included unfilled pauses, repeats, repairs, and filled pauses. Unfilled pauses included silences that could be associated with a word retrieval difficulty: silences within words, silences between simple modifiers and heads, between heads and simple complements, between compound verbs and compound noun phrases (e.g. "fall [pause] down"). Silences were not considered a pause if they occurred after a discourse marker (e.g. "well", "you know", "but", etc.), after reporting verbs (e.g. "I think"), or between phrases. Repeats included repetition of words that were exactly the same as what came before (e.g. "I went to the [pause] to the store"). Repairs included modifications in speech where what came after was meant to overwrite what came before (e.g. "I went to the shore [pause] to the store"). Filled pauses included "umm"s and "uhh"s. If a filled or an unfilled pause occurred as part of a repair or repeat strategy (e.g. "I went to the, uhh, to the store"), that was counted as a repair or repeat respectively.

Participants produced a total of 1905 disfluencies (Gestures visible: 988; Gestures not visible: 917). Speech disfluencies included 457 unfilled pauses, 232 repeats, 391 repairs, and 824 filled pauses. A clause was classified as Disfluency Present if it included at least one disfluency. 42 percent of the clauses (1327 out of 3174) had at least one disfluency associated with them. Coder 3 coded speech disfluencies for 10% of all stories and the intercoder agreement for whether a clause contained disfluencies or not was 93% (Cohen's κ $= .86, z = 15.3, p < .0001$).

Gesture coding

Analyses of the gestures in both of the visibility conditions were reported in Yap and colleagues (2018), but no analysis about the relationship between gestures and disfluencies were reported previously. Yap and colleagues (2018) recorded the gesture strokes for each clause, excluding self-adaptive actions. Gesture strokes were classified into different gesture types (i.e. beat, iconic, metaphoric, deictic or emblem) in two stages. In the first stage, using only the video with no audio, each gesture stroke was determined to be a beat or non-beat according to McNeill's (1992) beat filter. Gestures were classified as beats if they had (i) two movement phases, (ii) a relaxed handshape, and (iii) movement only within a single region of gesture-space. Gestures that included any features of other gesture types (e.g. deictics or iconics) were not classified as beats.

In the second stage, using both audio and video, the non-beat gestures were classified into iconics, metaphorics, deictics, or emblems, according to the gestures' forms and the accompanying speech. Participants produced a total of 3192 gestures (Gestures visible: 2012; Gestures not visible: 1180). Gestures included 446 iconics, 65 metaphorics, 2392 beats, 288 deictics and 1 emblem, according to McNeill's (1992) gesture categories.

Semantic and non-semantic gestures

We classified iconics, metaphorics, and beats into gestures that conveyed semantic meaning and thus contributed to the official business of presenting an utterance, and gestures that did not. Iconics and metaphorics were always categorized as semantic gestures since they convey semantic meaning. Beats, on the other hand, can either convey semantic meaning or not. Although beats have been traditionally thought to be devoid of semantic meaning (the direction of beat gestures are thought to not reflect anything about the content of speech, but are seen as random flicks of the hand, see for example McNeill, 1992), Yap and colleagues (2018), analyzing the same dataset we used here, showed that many beat gestures in fact reflect spatial semantics. For example, people's beat gestures were more likely to be upward when their speech implied upward motion (e.g. "my rocket went higher"), more frequently than expected by chance. Beat gestures were congruent with the story direction not only during literal spatial language (e.g. "my rocket went higher"), but also when participants used spatial metaphors for abstract motion (e.g. "my grades went higher"), and when they expressed the same abstract ideas without using any spatial language (e.g. "my grades got better").

We therefore categorized beats as congruent or non-congruent with the spatial semantics implied by the stories. To do so, we used Yap and colleagues' (2018) coding of the beat strokes according to their direction (upward, downward, leftward, rightward, other), using silent videos. For strokes with more than one directional component (e.g., down and left), the dominant direction was recorded. Congruent beats were considered semantically meaningful; non-congruent beats were considered non-meaningful.

Beat gestures that occurred during spatial clauses were classified as congruent if the direction of the beat gesture (e.g. upward) was the same with the direction implied by the spatial clause (e.g. upward) and as non-congruent if the direction of the beat gesture (e.g. upward) was different from the direction implied by the spatial clause (e.g. downward, leftward or rightward). Furthermore, beat gestures that occurred during non-spatial clauses ("I'm testing some new model rockets") were also classified as congruent if the direction of the beat gesture (e.g. upward) was the same with the direction implied by the whole story (e.g. upward) and as non-congruent if the direction of the beat gesture (e.g. upward) was different from the direction implied by the whole story (e.g. downward, leftward or rightward).

Overall, participants produced a total of 1356 semantic gestures including 446 iconics, 65 metaphorics, and 845 congruent-beats and a total of 1521 non-semantic gestures (i.e. non-congruent beats). 26 of the beat gestures were excluded, because either they could not be categorized as congruent or non-congruent as multiple spatial directions were implied by the accompanying clause, or the coder was unsure about coding the gesture as a beat.

A clause was classified as Gesture Present if it included at least one gesture (iconic, metaphoric, beat or deictic). 49 percent of the clauses (1563 out of 3174) had at least one gesture associated with them. A clause was classified as Non-semantic Beat Gesture Present if it included non-semantic gestures only (non-congruent beats, and not iconics, metaphorics or congruent beats). A clause was classified as Semantic Iconic Gesture Present if it included iconic gestures only (and no other gestures); and so on (Semantic Metaphoric Gesture Present; Semantic Beat Gesture Present) for each of the semantic gesture types.

3.2.5 Analyses

We conducted all analyses by fitting generalized linear mixed-effect models, using R (R Core Team, 2020), the glmer() function in the lme4 library (Bates et al., 2015). We used a "maximal" random effect structure justified by our design (Barr et al., 2013). We treated Subject (N = 53) and Story (N = 15) as random effects, including random intercepts for both in analyses, since our outcome variables (e.g. Gesture Presence) are likely to vary across different subjects and stories. Subjects are likely to vary idiosyncratically also in their sensitivity to our fixed factor, Disfluency Presence. Therefore, we also included random slopes for our within-subject fixed factor (i.e. Disfluency Presence), allowing subjects' likelihood to gesture to vary differentially based on our fixed factor (e.g., subjects could be affected differently by Disfluency Presence in their production of gestures). We used likelihood ratio tests (LRTs) to test for the omnibus interaction effects. We estimated group means and performed our planned contrasts using the emmeans() function in the emmeans package (Lenth et al., 2018) in R. We reported Odds Ratios (OR) in the response scale (probability) as a measure of effect size for the simple effects.

3.3 Results

3.3.1 *Effect of disfluency and visibility condition on gesture production*

To compare Gesture Presence across experimental conditions for each clause (N = 3174), we used mixed effects logistic regressions with Gesture Present and No gesture as the binary outcomes for a clause.[1]

We first tested whether people were more likely to gesture when they were disfluent than when they were fluent during face-to-face conversation, that is, when their gestures were visible. Conceptually replicating previous work, we showed that participants whose gestures were visible were indeed more likely to gesture during disfluent clauses than fluent clauses ($z = 4.25, p < .001, OR = 2.09, 95\% CI[1.34, 3.25]$; see the fluent and disfluent columns on the left in Figure 3.1).

Next, we tested whether people gestured more during disfluent clauses only (or more so) when their gestures were visible to the listener, but not (or less so) when their gestures were not visible. To do so, we tested whether visibility (Gesture Visible, Gesture Not Visible) changed the effect of disfluency (Disfluent Clause, Fluent Clause) on gesture production (Gesture Present, No Gesture). As predicted, results showed a significant two-way interaction between visibility condition and disfluency on gesture production: People were more likely to gesture during disfluent clauses than fluent clauses only when their gestures were visible ($\chi^2(1) = 4.10, p = .043$). For participants whose gestures were not visible, there was no significant effect of disfluency on gesture production ($z = 1.27, p = .579, OR = 1.25, 95\% CI[0.79, 1.98]$, see the fluent and disfluent columns on the right in Figure 3.1).

---

[1] R syntax for the omnibus interaction model: gesture presence ~ disfluency presence * gesture visibility + (1 + disfluency presence | subject) + (1 | story)

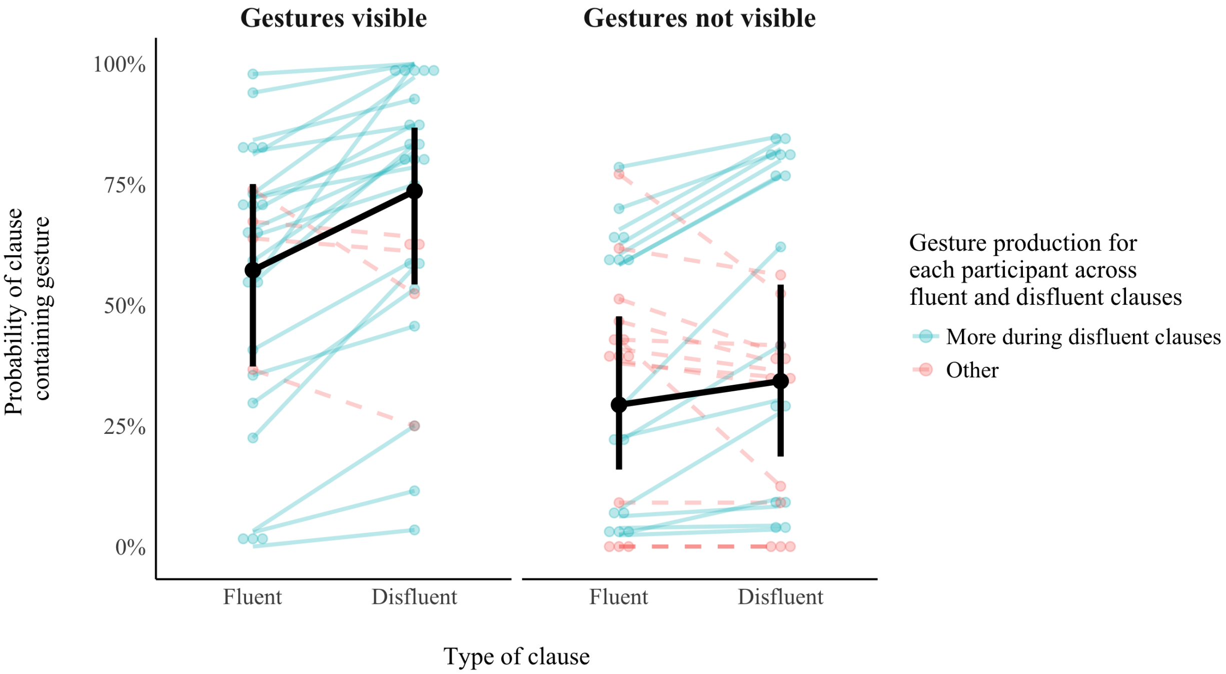

Figure 3.1: Results showing the probability of clauses containing gestures during fluent and disfluent speech (# of clauses that contain gestures / # of all clauses), for participants whose gestures were visible (left) and participants whose gestures were not visible to their interlocutor (right). Black points represent the estimated group means and black error bars around these points represent the estimated 95% confidence intervals (asymp.LCL and asymp.UCL) for the mixed model, calculated using the emmeans() function in the emmeans package in R. Colored lines represent the data summary with the means for individual participants (Solid blue = Probability of gesture production greater during disfluent clauses, Dashed red = Probability of gesture production not greater during disfluent clauses).

3.3.2 Effect of disfluency and visibility on non-semantic and semantic gestures

Do all gestures serve as pragmatic signals commenting on the problems with the act of speaking, or only some gestures? Non-semantic gestures (beats that do not represent speech) are not signals in the primary track – they do not contribute to the official business of presenting an utterance. Non-semantic gestures should get their semiotic value from being signals in the secondary track, serving as pragmatic signals that comment on the act of speaking. If so, then non-semantic gestures should be more likely to be produced during disfluent clauses than fluent clauses only (or more so) when gestures are visible, and not (or less so) when they are not visible.

19 percent of the clauses (608 out of 3174) had only non-semantic gestures associated with them. To compare Non-semantic Gesture Presence across experimental conditions for each clause (N = 2318 containing 608 clauses with non-semantic gestures only and 1710 clauses with no gestures), we used mixed effects logistic regressions with Non-semantic Gesture Present and No Gesture Present as the binary outcomes for a clause.[2] We excluded clauses that contained semantic gestures for this analysis (856 clauses, total, containing 559 clauses with semantic gestures only; 297 clauses with both semantic and non-semantic gestures).

We first tested whether people were more likely to produce non-semantic gestures than to produce no gestures when they were disfluent, compared to when they were fluent during face-to-face conversation (i.e., when their gestures were visible). We showed that participants whose gestures were visible were indeed more likely to produce non-semantic gestures than to produce no gestures during disfluent clauses, compared to fluent clauses ($z = 4.36, p < .0001, OR = 2.27, 95\% CI[1.57, 3.29]$; see the fluent and disfluent columns on the left in Figure

---

[2] R syntax for the omnibus interaction model: non-semantic gesture presence ∼ disfluency presence * gesture visibility + (1 + disfluency presence | subject) + (1 | story). Note that the data in the model did not include clauses that have semantic gestures – neither clauses that only have semantic gestures nor clauses with both semantic and non-semantic gestures.

3.2).

Next, we tested whether the link between producing non-semantic gestures and disfluencies was found only (or more so) when gestures were visible to the listener. To do so, we tested whether visibility (Gesture Visible, Gesture Not Visible) changed the effect of disfluency (Disfluent Clause, Fluent Clause) on non-semantic gesture production (Non-semantic Gesture Present, No Gesture Present). Results showed a significant two-way interaction between visibility and disfluency on non-semantic gesture production: People were more likely to produce non-semantic gestures than to produce no gestures during disfluent clauses, compared to fluent clauses, but only when their gestures were visible ($\chi^2(1) = 5.18, p = .023$). For participants whose gestures were not visible, there was no significant effect of disfluency on non-semantic gesture production ($z = 0.88, p = .377, OR = 1.20, 95\%CI[0.80, 1.82]$, see the fluent and disfluent columns on the right in Figure 3.2).

We showed clear evidence that non-semantic gestures serve as pragmatic signals commenting on problems with the act of speaking. Next we turned to semantic gestures – iconics, metaphorics and beats that represent speech. First, is each type of semantic gesture also more likely to be produced when people are disfluent, compared to when they are fluent during face-to-face conversation, that is when gestures are visible? Second, if there is a link between production of each type of semantic gesture and disfluencies, is this link pragmatically motivated? Does each type of semantic gesture also comment on problems with the act of speaking? Semantic gestures could be signals in two tracks simultaneously, both contributing to the official business and also commenting on deviations from the official business. If this is true, semantic gestures should be more likely to be produced during disfluent clauses, compared to fluent clauses, when gestures are visible but not (or less so) when they are not visible.

Alternatively, only non-semantic gestures may take up the role of pragmatic signals commenting on problems with the act of speaking. If this is true, semantic gestures may not be

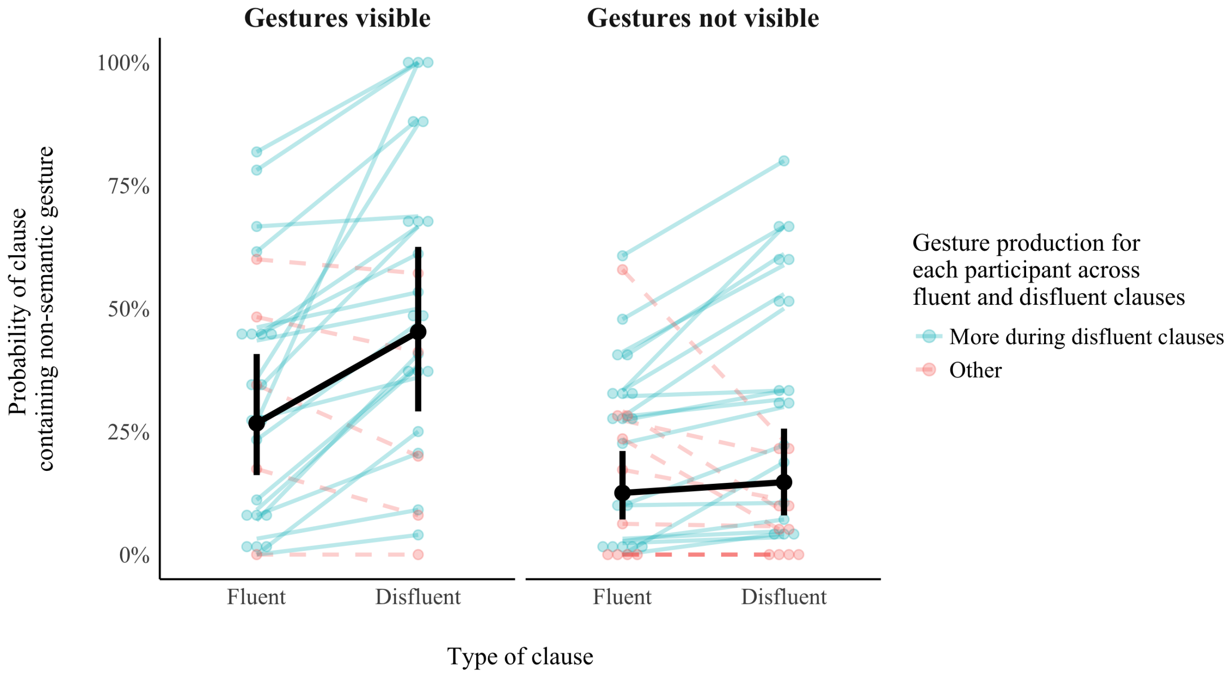

Figure 3.2: Results showing the probability of clauses containing non-semantic gestures (# of clauses that contain non-semantic gestures only / (# of clauses that contain non-semantic gestures only + # of clauses with no gestures) for fluent and disfluent clauses, for participants whose gestures were visible (left) and participants whose gestures were not visible to their interlocutor (right). Black points represent the estimated group means and black error bars around these points represent the estimated 95% confidence intervals (asymp.LCL and asymp.UCL) for the mixed model, calculated using the emmeans() function in the emmeans package in R. Colored lines represent the data summary with the means for individual participants (Solid blue = Probability of non-semantic gesture production greater during disfluent clauses, Dashed red = Probability of non-semantic gesture production not greater during disfluent clauses).

more likely to be produced during disfluent clauses, compared to fluent clauses – even when gestures are visible. Alternatively, there might be a link between semantic gestures and disfluencies, but this link may not be pragmatically motivated, so the visibility of gesture may not matter. For example, if semantic gestures are reliable symptoms of speech difficulties, people should be likely to gesture more during disfluent speech whether or not their gestures are visible.

To test these hypotheses, we looked at whether people were more likely to gesture than not gesture based on disfluency and visibility for each of the semantic gesture types. 18 percent of the clauses (559 out of 3174) had only semantic gestures associated with them – containing 189 clauses with iconic gestures only, 35 clauses with metaphoric gestures only, 286 gestures with congruent beats only, and 49 clauses with some mixture of these semantic gesture types.

To compare Iconic Gesture Presence across experimental conditions for each clause (N = 1899 containing 189 clauses with iconic gestures only and 1710 clauses with no gestures), we used mixed effects logistic regressions with Iconic Gesture Present and No Gesture as the binary outcomes for a clause. To compare Metaphoric Gesture Presence across experimental conditions for each clause (N = 1745 containing 35 clauses with metaphoric gestures only and 1710 clauses with no gestures), we used mixed effects logistic regressions with Metaphoric Gesture Present and No Gesture as the binary outcomes for a clause. To compare Semantic Beat Presence across experimental conditions for each clause (N = 1996 containing 286 clauses with congruent beat gestures only and 1710 clauses with no gestures), we used mixed effects logistic regressions with Semantic Beat Present and No Gesture as the binary outcomes for a clause.[3]

---

[3] R syntax for the omnibus interaction model for iconic gesture presence as an example: iconic gesture presence ∼ disfluency presence * gesture visibility + (1 + disfluency presence | subject) + (1 | story). Note that the data in the model did not include clauses that have non-semantic gestures or any semantic gesture type that is not iconic gestures. The random effect structure was the same for congruent beat presence. For metaphoric gesture presence, the model with the full random effect structure did not converge, and we used an intercept only model instead (i.e. (1 | subject) + (1 | story)).

We first tested whether people were more likely to produce each of the three semantic gesture types when they were disfluent than when they were fluent during face-to-face conversation, that is, when their gestures were visible. Contrary to previous claims, there was no statistically significant evidence that people were more likely to produce iconics, metaphorics or congruent beats during clauses with disfluencies than during clauses with no disfluencies when their gestures were visible (iconics: $z = 1.62, p = .104, OR = 1.64, 95\% CI[0.90, 3.00]$; metaphorics: $z = 1.25, p = .213, OR = 1.90, 95\% CI[0.70, 5.19]$; congruent beats: $z = 1.41, p = .157, OR = 1.47, 95\% CI[0.86, 2.50]$; see the fluent and disfluent columns on the left in Figure 3.3). Since people did not seem more likely to produce iconics, metaphorics, or congruent beats when they were disfluent compared to when they were fluent, we did not expect (and did not find) modulation of this behavior by the visibility condition.

3.4 Discussion

People are more likely to gesture when their speech is disfluent, compared to when their speech is fluent (Akhavan et al., 2016; Butterworth & Beattie, 1978; Ragsdale & Silvia, 1982). Here we explore why. We found that people do indeed gesture more when their speech is disfluent, but only when those gestures are visible to the listener in face-to-face conversation. Our findings thus suggest that gestures produced during disfluent speech serve as pragmatic signals, commenting to the listener on the speaker's difficulty presenting an utterance. Overall, our results suggest that gesturing when disfluent is not merely a symptom of speech difficulties. Rather, gesturing when disfluent serves as a pragmatic signal to the listener commenting on the speaker's problems with speaking.

3.4.1 Gestures are not mere symptoms of speech problems

It is clear that some gestures serve as communicative signals, meant to be seen by a listener (Bavelas et al., 1992). However, gestures that occur when speaking is difficult have been

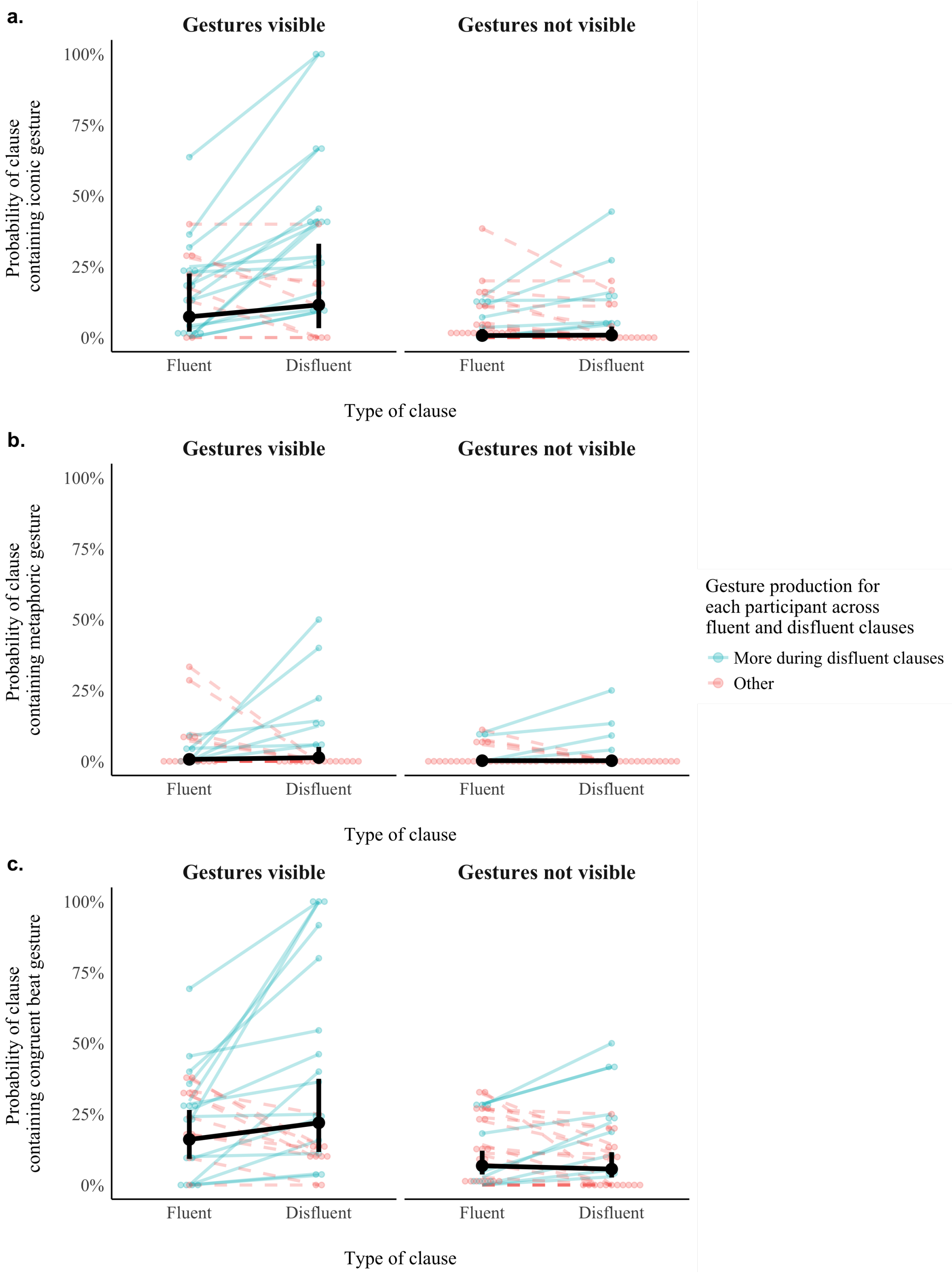

Figure 3.3: Results showing the probability of clauses containing semantic gestures, a. iconics, b. metaphorics, and c. congruent beats, for fluent and disfluent clauses, for participants whose gestures were visible (left) and participants whose gestures were not visible to their interlocutor (right).

Figure 3.3, continued: For each semantic gesture type, the probability of clauses containing that gesture type was calculated according to the same formula: # of clauses that contain that semantic gesture type only (e.g. iconic gestures only) / (# of clauses that contain that semantic gesture type only + # of clauses with no gestures). Black points represent the estimated group means and black error bars around these points represent the estimated 95% confidence intervals (asymp.LCL and asymp.UCL) for the mixed model, calculated using the emmeans() function in the emmeans package in R. Colored lines represent the data summary with the means for individual participants (Solid blue = Probability of semantic gesture production greater during disfluent clauses, Dashed red = Probability of semantic gesture production not greater during disfluent clauses).

widely interpreted as evidence that many gestures are primarily produced to meet speaker needs. Gestures have been hypothesized to grow out of difficulty with formulating what to say, a hypothesis about a mechanism for gesturing, and to help speech production, a hypothesis about a function of gesturing (Krauss & Hadar, 1999). The present data are not compatible with either hypothesis, however, for two reasons. First, the gestures that have been hypothesized to grow out of speech problems, semantic gestures (i.e. iconics, metaphorics, or congruent beats), were not even more likely to occur when people encountered speech problems – that is, when they were disfluent. Second, non-semantic gestures (i.e. non-congruent beats), gestures that were more likely to occur when people encountered speech problems, increased only when speakers encountered production difficulties *and* their gestures were visible. When non-semantic gestures were not visible, and therefore could not serve as a pragmatic signal to the listener, they were *not* more likely to occur with disfluent clauses. If gestures were simply symptoms of speech difficulties, they should co-occur with speech problems regardless of whether they are visible or not.

The finding that gestures are more likely to occur with disfluent speech also has been taken as evidence that gestures might facilitate speech production. However, given that this behavior, when it occurs, seems to be pragmatically motivated, the correlation between gestures and disfluencies does not need to be interpreted as indicating that gestures are produced to facilitate speech production. The need to reinterpret any correlation between gesture rate and disfluency rate is underscored by a previous study showing that there is, in fact, no reliable evidence that preventing gestures impairs speaking (Study 1; see also Kısa et al., 2021).

3.4.2 Which gestures serve as pragmatic signals during disfluent speech?

In principle, all gestures that are produced when people speak could serve as pragmatic signals commenting on problems with speaking. Accordingly, both the gestures that contribute to the official business of presenting an utterance, and the gestures that do not contribute to the official business, might serve as pragmatic signals commenting on deviations from the official business.

Alternatively, different gestures may be produced for different reasons. In fact, David McNeill (2008) suggested a division of labor between different gesture types. He suggested that the semiotic value of gestures that represent speech comes from their semantic meaning – that is, from their contribution to the official business of presenting an utterance. In contrast, the semiotic value of beats, gestures that have been traditionally thought to be devoid of semantic meaning, comes from their pragmatic meaning as a commentary on the act of speaking.

We tested these hypotheses by looking at the types of gestures that were more likely to be produced during disfluent speech *and* were visible. Our results provided clear evidence of a pragmatic motivation in the case of non-semantic gestures. We showed that, compared with fluent clauses, disfluent clauses were approximately two times more likely to have a non-semantic gesture than to have no gesture – but *only* when those gestures were visible. This pattern suggests that producing non-semantic gestures during disfluent speech is pragmatically motivated: Non-semantic gestures serve as pragmatic signals commenting on problems with speaking.

In contrast, for semantic gestures, we could not test whether their link to disfluent speech was pragmatically motivated, because we failed to find clear evidence for that link: Contrary to earlier claims, we failed to find statistically significant evidence that semantic gestures are more likely to be produced during disfluent, compared to fluent speech – even when gestures were visible. Because this is a null result, it remains possible that semantic gestures are linked to disfluent speech, but we simply failed to find evidence for it. Even though we had, overall, a large number of observations in this study (3174 clauses), we had comparatively fewer observations for each of the semantic gesture types (189 clauses with iconics only, 35 clauses with metaphorics only, and 286 clauses with congruent beats only) than for the non-semantic gestures (608 clauses with non-congruent beats only). Our test for non-semantic beat gestures was thus higher powered than for semantic gestures. If it does turn out that there is a link between semantic gestures and disfluencies that we failed to detect, it remains possible that this link is also pragmatically motivated by the need to signal problems with speech production – as it is for non-semantic gestures. In other words, our results leave open the possibility that both semantic and non-semantic gestures are more likely to be produced during disfluent speech *and* that this behavior is pragmatically motivated by the need to signal problems with speech production. However, our results provide clear evidence only for non-semantic gestures.

3.4.3 Pragmatic motivations for gesturing

We know that gestures serve many pragmatic functions, commenting on the act of speaking in different ways. Gestures highlight what is important with respect to the larger discourse by commenting on the addition of new characters and information or departing briefly from the narrative (Alibali, Heath & Myers, 2001; McNeill, 1992, 2008). Gestures also comment on the act of conversing with another person by asking for, or citing, the listener's contribution and by managing turn-taking during dialogue (Bavelas et al., 2015; Beattie & Aboudan 1994). Here, we introduced a novel pragmatic motivation for gesturing. Gestures serve as a pragmatic signal by commenting on drifting away from, and having problems with, the official business of presenting an utterance.

In line with this pragmatic account of gesturing during disfluent speech, Beattie & Aboudan (1994) showed that people are more likely to gesture when they pause than when they utter a word, only when they are in a conversation, but not when they produce a monologue (see Nobe, 2000, for a replication). So gestures can signal the speaker's intent to continue speaking despite the current pause in speech, holding the floor to prevent the listener from interrupting their turn (Butterworth & Hadar, 1989; Duncan, 1972) – one way in which speakers can comment on having problems with speaking.

Pausing at potential turn boundaries, however, is only one among many ways that speakers can deviate from the official business of presenting an utterance. In this chapter, we focused on a case where speakers were engaged in a monologue with minimal need for holding the floor and focused on types of disfluencies that are less likely to license turn interruption. We focused on pauses in the middle of a phrase where a potential turn interruption may be a minimal problem, and on disfluencies other than pauses (repairs, repeats and fillers) that should not license turn interruption. Yet, these deviations from fluent speech still interrupt the official business of presenting an utterance and speakers may need to signal this deviation. Deviations from an utterance plan are also deviations from the conversation timing plan – and commenting on speech problems can give the listener information about the speaker's production plan and thus ensure successful coordination in conversation timing (Holler & Levinson, 2019). Additionally, deviations from presenting an utterance are also deviations from the role a speaker commits to by being in a joint action of conversing. Speakers commit to presenting an utterance when they have the turn and listeners expect them to commit to this role and give an account of any deviation from this role (Clark, 1996). For these reasons, gesturing might not only have the specific motivation to ensure not losing your turn when your speech stops at a potential turn boundary, they might also have a more general motivation to comment on problems with speaking whenever your speech is drifting away from the official business of presenting an utterance.

In this chapter, we have provided evidence for a novel pragmatic motivation for gesturing: Gestures comment on problems with speaking when they are visible to a listener. This commentary can come in many forms, and may arise during different kinds of speech difficulties. Further analyses of which kinds of gestures tend to occur during which kinds of speech disfluencies could reveal the specific ways in which gestures comment on problems with speaking.

The communicative motivations for gesturing when speaking is difficult are likely to be more wide-spread than the cases we discuss here. People gesture not only when they are disfluent, but also when speaking is difficult and no speech disfluencies are produced. For example, people are more likely to gesture when they are about to utter a low probability word – words that are likely hard to retrieve from the lexicon for the speaker (Beattie & Shovelton, 2000). Gestures may be produced with communicative motivations in these situations as well, rather than being mere symptoms of speaking difficulty. In the next chapter, we test a communicative motivation for why people may be more likely to gesture before a low probability word: Because gestures may signal and depict hard to predict words for the listener.

CHAPTER 4

A COMMUNICATIVE ACCOUNT OF WHY SPEAKERS GESTURE TO DEPICT LOW PROBABILITY WORDS

4.1 Introduction

When people talk, they gesture. And some gestures depict what the speaker says – they convey semantic meaning. These gestures do not depict just anything though. Semantic gestures tend to depict words that are unexpected given its prior linguistic context – that is words that have a lower transitional probability (Beattie & Shovelton, 2000). One way to calculate transitional probability for words, say for a sentence, is to have people guess each successive word in that sentence from its preceding context and calculate the empirical probability of correct guesses among all the guesses made (using the guessing procedure in Shannon, 1951). Calculating transitional probabilities in this way, Beattie & Shovelton (2000) found that the words that gestures depicted (i.e. gestures' lexical affiliates) had lower transitional probabilities than other words that are matched to lexical affiliates in terms of their grammatical category – that is other nouns and verbs in the sentences.

Why do gestures depict words that have a low probability of continuing that utterance? According to one proposal, people may be more likely to gesture to depict a low probability word, because these words are difficult for the speaker to access from the lexicon; producing a gesture can help the speaker access that word more easily (Hadar & Butterworth, 1997; Krauss & Hadar, 1999). However, as I argued in Chapter 2, gestures do not appear to help speech production by helping people access the right words. But why then is there a tight relationship between gestures and low probability words? In this chapter, I propose that this relationship may be communicatively motivated.

Speakers may be more likely to gesture before a low probability word than a high probability word because the gestures may serve as a pragmatic signal to the listener that a relatively less predictable word is coming up. Additionally, the form of the gesture may depict information for the listener, so that they can more easily predict and process relatively unpredictable, highly informative, parts of the utterance.

People prefer to avoid high information peaks when they speak, because a more uniform information distribution across an utterance may make it easier for a listener to predict and process that utterance (Aylett & Turk, 2004; Levy & Jaeger, 2007). Parts of speech that are less frequent, less familiar, or less predictable are highly informative, and people prefer to spread these highly informative parts of speech out to have a more uniform information distribution. For example, people elongate less frequent words (e.g. 'thyme' is pronounced longer than 'time'; Gahl, 2008) or add extra words to phrases that are less predictable (e.g. using ‘that’ in ‘the family that you cook for’ but not in ‘the family you live with’; Levy & Jaeger, 2007).

Gestures could also be used to achieve a more uniform information distribution. Gestures tend to precede the words that they depict, even up to by 4 seconds (Morrel-Samuels & Krauss, 1992; ter Bekke et al., 2020). Speakers may be more likely to produce a gesture before a highly informative, hard to predict word to make it more predictable for the listener. Gestures that depict may do this in at least two ways. One, the presence of a gesture may signal that something informative is coming up – so that the listener can guide their attention accordingly and more easily process the utterance. Second, the form of gesture may convey part of the semantics of the upcoming hard to predict word – making it easier for the listener to predict and process the word (Holler & Levinson, 2019). In fact, there is work showing that semantic gestures help listeners disambiguate words and better process less predictable words (Holle & Gunter, 2007; Zhang et al., 2021).

In this study, our aim was to first conceptually replicate the findings that gestures are affiliated with low probability words during face-to-face conversation (Beattie & Shovelton, 2000). We then ask whether this relationship between gestures and word probability is communicatively motivated. If semantic gestures are produced with a communicative motivation to signal and depict unpredictable words for the listener, then gesture's lexical affiliates should have a lower probability of continuing the utterance (i.e. transitional probability), compared to other nouns and verbs – and this should be true only (or more so) when the listener can see the gestures.

Alternatively, if gestures are not produced with a communicative motivation to signal and depict unpredictable words for the listener, then the visibility of the gestures should not matter. That is, if gesture production is a mere symptom of difficulty accessing low probability words, driven solely by speaker needs, gestures should depict words with lower transitional probability regardless of whether the listener can and cannot see the gestures.

4.2 Methods

The design and analysis plan for the study were pre-registered.

4.2.1 Source corpus

We selected a sample of sentences from the corpus of gesture and speech we used in Study 2 with two visibility conditions: gestures were visible and gestures not visible. In the gestures visible condition, participants were seated facing one another across a table. In the gestures not visible condition, the listener was blindfolded and the participants were separated by an opaque barrier on the table's surface occluding gesture space. This corpus contained a total of 3534 spoken clauses (Gestures visible: 1936; Gestures not visible: 1598) with 446 iconic and 65 metaphoric gestures.

4.2.2 Target sentences

From this corpus of story retellings, we identified parts of speech to be used for guessing, to calculate the transitional probabilities. We filtered sentences that have only one iconic or metaphoric gesture associated with them, following Beattie & Shovelton (2000). This procedure resulted in 33 target sentences. The number of sentences that have only one depicting gesture, 33, is low compared to the total number of iconic and metaphoric gestures in the source corpus, 511, since most sentences tend to have more than one iconic or metaphoric gesture. However, here we work with a higher number of sentences than what has been done previously: Beattie & Sholveton (2000) analyzed 12 sentences only.

This filtering decision was motivated by two reasons. One, with the aim of conceptually replicating and extending Beattie & Shovelton (2000)'s findings we tried to keep our design similar to theirs. Second, focusing on sentences, rather than phrases or clauses, that have one gesture associated with them allowed us to have a strong test of the idea that lexical affiliates of gestures have lower probability than other nouns and verbs: There are many other nouns and verbs within a sentence to which one can compare the transitional probability of lexical affiliates to, whereas there are not many other nouns and verbs within a clause or phrase.

We further excluded a total of 6 sentences since (i) 2 of them could not be understood in isolation, (ii) for 3 of them the video quality was low that prevented gesture coding, and (iii) for 1 of them the gesture did not terminate within the target sentence. After the exclusions, we ended up with 27 target sentences (Gestures visible: 13; Gestures not visible: 14).

4.2.3 Guessing

The 27 target sentences included a total of 404 word tokens (a mean of 15 words per sentence). We created 3 sets of target sentences, with 9 sentences in each, such that no two target sentences came from the same source sentence in the original story scripts – to avoid individual guessers seeing a retelling of the same source sentence twice. We recruited 60

University of Chicago undergraduates in total to guess all the word tokens – so, each set of target sentences were seen by 20 guessers.

We used Qualtrics to randomly assign guessers to one of the three sentence sets and presented 9 target sentences to individual guessers in random order. We presented the target sentences with transcriptions of the preceding context, if the story that they were part of included sentences that came before the target sentence. The guessers were asked to guess the first and then each successive word in each target sentence, continuing to the last, until the sentence was complete (based on the Shannon, 1951 guessing procedure, used by Beattie & Shovelton, 2000). Each guesser was given two guesses for each successive word. After they made two guesses, they were given the correct word and asked to continue to the next word. Each sentence set had a comparable number of total word tokens guessed (134 word tokens in Set 1, 125 in Set 2 and 145 in Set 3).

For guessing purposes, we used the citation forms of the sentences rather than the raw transcripts, following Beattie & Shovelton (2000). To do so, we deleted disfluencies (e.g. ‘um’) and replaced contractions with separate words so that each word can be guessed separately (e.g. “I’m” as “I am”).

A total of 16160 word guesses were made (40 guesses for each of the 404 word tokens), and 438 guesses were excluded where 2 were non-word entries, 164 were second guesses for the same word token repeating the first guess, and 270 were incorrect guesses made by two participants who misunderstood the instructions (their second guess was not a guess for the same word, but a guess for what could come after their first guess, for example entering “am” as their first guess and “writing” as their second guess for what word continues the utterance “I . . . ”, rather than making two guesses for the same word such as “am” and “do”).

4.2.4 Transitional probability

We calculated the transitional probability score for each word token based on the frequency of correct guesses divided by the sum of the different words (i.e. types) suggested by the guessers, following Beattie & Shovelton (2000).

4.2.5 Word category

For each word token, we determined whether it is a noun or verb, or whether it is any other word category.

Additionally, we recruited 11 University of Chicago undergraduates to identify the lexical affiliates in each of the target sentences on Zoom. We prepared videos that included only the target sentences, rather than the entire story. For each of the 27 videos, participants first read the citation form of the target sentence on Qualtrics, then the experimenter played the video three times and then participants were asked to select the place in the sentence where the meaning of a word is related to the meaning of the gesture, using whatever criterion seems appropriate to them. The participants were allowed to select only one word for each sentence.

A lexical affiliate was identified for each sentence if the same word was selected by 8 out of 11 people. Out of the 27 sentences, 16 sentences had a clear lexical affiliate (Gestures visible: 8; Gestures not visible: 8). These 16 sentences were used for further coding and analyses. The final list of sentences included 256 word tokens containing 16 lexical affiliates and 78 other nouns and verbs. Following Beattie & Shovelton (2000), we compared lexical affiliates' transitional probability to other words that are mostly matched in terms of grammatical category, other nouns and verbs.

4.2.6 Analyses

We conducted our analyses by fitting linear models, using R (R Core Team, 2020), the glmer() function in the lme4 library (Bates et al., 2015) and the lm() function in the stats library. We used "maximal" random effect structures justified by our design (Barr et al., 2013). We used likelihood ratio tests (LRTs) to test for our planned contrasts using the anova() function.

4.3 Results

To compare Transitional Probability for each word (N = 94) across experimental conditions, we used mixed effects Poisson regressions with an offset term. Transitional probability is a rate, rather than a count, with the number of correct guesses divided by the number of the different word types suggested by the guessers. Therefore, to model a rate variable with Poisson regressions, we used the number of correct guesses as the outcome variable for each word and we used the number of different word types guessed for each word as an offset term into the model (Agresti, 2003).

We treated Target Sentence (N = 16) as a random effect, including random intercepts in analyses, since our outcome variable, transitional probability of a word, is likely to vary across different target sentences. We also included random slopes for our fixed effect of interest that is within-sentence, Word Category since the effect of word category (whether a word is a lexical affiliate or other noun or verb) on the transitional probability of words are also likely to vary across different target sentences (e.g. the difference in transitional probability of lexical affiliates versus other nouns and verbs might be higher for some sentences).[1]

---

[1] Example syntax for the omnibus interaction model: # of correct guesses ∼ lexical affiliate status * gesture visibility + offset(log(# of word types guessed)) + (1 + lexical affiliate status | sentence)

4.3.1 Do lexical affiliates of gestures have a lower transitional probability than other nouns and verbs when gestures are visible?

First, we tested whether gestures depict words in a sentence that have lower transitional probability than other nouns and verbs during normal face-to-face conversation when gestures are visible. Conceptually replicating previous work (Beattie & Shovelton, 2000), we found that lexical affiliates have a lower transitional probability than other nouns and verbs when gestures are visible ($\chi^2(1) = 5.40, p = .020$; see Figure 4.1 gestures visible).

4.3.2 Do lexical affiliates of gestures have a lower transitional probability than other nouns and verbs, only (or more so) when gestures are visible?

Next, we tested whether gestures do not, or are less likely to, depict words with lower transitional probability (than other nouns and verbs) when they are not visible. In line with the pragmatic account we proposed, when gestures were not visible, we failed to find a statistically significant effect that lexical affiliates have a lower transitional probability compared to other nouns and verbs ($\chi^2(1) = 0.15, p = .695$; see Figure 4.1 gestures not visible).

We further tested the two-way interaction between word category and gesture visibility to see whether the presence of an effect of word category on transitional probability for gestures visible condition was statistically different from its absence in the gestures not visible condition. We failed to find statistically significant evidence (at $p < .05$), but showed weak evidence that gestures' lexical affiliates had a lower transitional probability compared to other nouns or verbs selectively (or more so) when gestures were visible ($\chi^2(1) = 2.62, p = .106$).[2]

---

[2] For comparison with Beattie & Shovelton (2000) that we aimed to conceptually replicate here, we also tested whether lexical affiliates had lower transitional probability than other words (i.e. not only other nouns and verbs but any other word category) when gestures were visible. Conceptually replicating Beattie

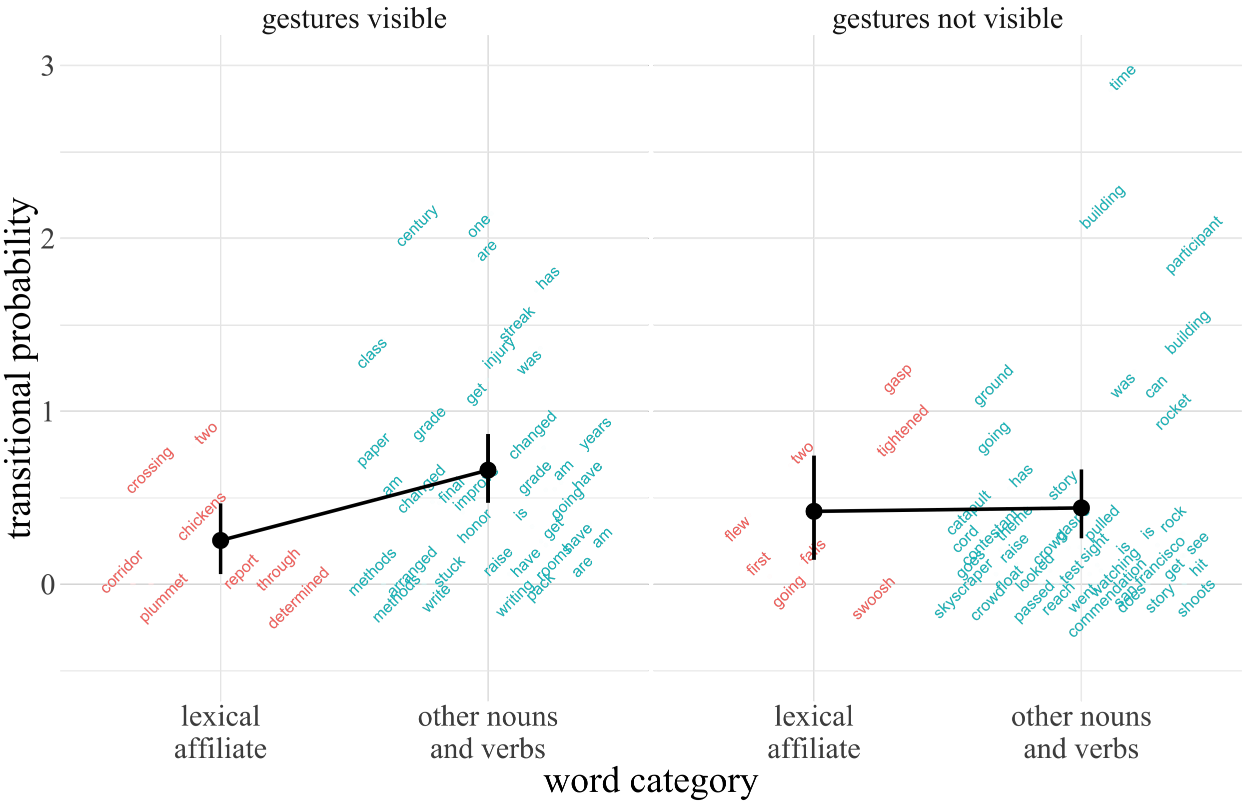

Figure 4.1: Results showing the transitional probability of lexical affiliates and other nouns and verbs, for participants whose gestures were visible (left) and participants whose gestures were not visible to their interlocutor (right). Black points represent the group means and black error bars around these points represent the bootstrapped 95% confidence intervals. Colored text indicate individual data points, that is words (Red = the word is a lexical affiliate, Blue = the word is a non-lexical affiliate noun or verb).

4.4 Discussion

Do gestures depict low probability words and, if so, why? Here, we showed gestures' lexical affiliates had a lower transitional probability than other nouns and verbs during face-to-face conversation, conceptually replicating Beattie & Shovelton (2000). Next, we showed that gestures' lexical affiliates did *not* have a lower transitional probability than other nouns and verbs when gestures were not visible, suggesting that semantic gestures' tendency to depict low probability words is communicatively motivated. Semantic gestures are not mere symptoms of difficulty accessing low probability words (a mechanism underlying gesture), instead, their production before low probability words is communicatively motivated: Semantic gestures depict low probability words because they signal and depict upcoming, hard to predict words for the listener.

The communicative account we proposed here predicted three effects: (i) when gestures are visible, lexical affiliates should have a lower probability than other nouns and verbs; (ii) when gestures are not visible, there should be either a weaker difference or no difference between the transitional probability of lexical affiliates and other nouns and verbs; and (iii) there should be an interaction between word category and gesture visibility on transitional probability. Overall, we found the pattern of results predicted by the communicative account: we found the predicted results for the first two effects and marginal evidence for the third prediction.

The marginal statistical evidence for the interaction may be due to low power. With the aim of conceptually replicating Beattie & Shovelton (2000), we followed their procedure

---

& Shovelton (2000), we found that lexical affiliates had a lower transitional probability than other words when gestures were visible ($\chi^2(1) = 5.70, p = .017$). However, as Beattie & Shovelton (2000) also argued, comparing the transitional probability of lexical affiliates to all other words is not a valid comparison to make (p. 485). This is because lexical affiliates tend to be content words, mostly nouns and verbs, and content words have lower transitional probability than all other words that also include non-content words. So, when one compares the transitional probability of lexical affiliates to all other words, lexical affiliates might have lower probability simply because nouns and verbs have lower probability compared to all other words. Therefore, comparing lexical affiliates' transitional probability to other words that are matched in terms of grammatical category, other nouns and verbs, is a more valid comparison to make.

in selecting gestures and lexical affiliates, filtering sentences from a corpus of gesture and speech that had only one iconic or metaphoric gesture. This decision resulted in a very few number of sentences, 8 per visibility condition, even in a large corpus like ours (over 3000 spoken clauses that likely make up around 1000 sentences), since sentences, if they do have a gesture, tend to have more than one gesture – it is rare to find a sentence with a single gesture. This was a great approach to do a strong test of the idea that lexical affiliates of gestures have lower probability than other nouns and verbs – since sentences often include multiple clauses and there are many other nouns and verbs within a sentence to which one can compare the transitional probability of lexical affiliates to. In future work, we will use clauses rather than sentences as the unit of analysis – such that we will include clauses that have a single gesture and have people identify a lexical affiliate for that gesture within the clause. This should result in a larger number of observations and thus a higher-powered test of our hypothesis.

4.4.1 Semantic gestures are not symptoms of word retrieval problems

We know that semantic gestures are, in part, designed for the listener, produced with a communicative motivation (Alibali et al. 2001; Holler & Bavelas, 2017). However, semantic gestures do not communicate about just anything – they tend to depict low probability words that are likely hard to retrieve for the speaker. This could be interpreted as evidence for the hypotheses that semantic gestures grow out of difficulty with formulating what to say, failing to find a word (a mechanism underlying gesture) and producing a gesture can activate semantic features that make it easier to retrieve that sought out word (a function that gesture serves; Krauss & Hadar, 1999). The present data are not compatible with these hypotheses, however, since gestures were not affiliated with low probability words when they cannot be seen by a listener. If gestures were simply symptoms of word finding difficulty, produced with the goal to resolve those difficulties, they should co-occur with low probability words regardless of whether they are visible or not.

4.4.2 Gestures signal and depict hard to predict words for the listener

Gestures depict low probability words that are hard to predict for the listener. Here, we showed that speakers not only modulate their speech (Gahl, 2008; Levy & Jaeger, 2007) to distribute highly informative arguments over a greater amount of time, but they also change what their gestures depict to avoid high information peaks for their listeners. This modulation may serve many functions for the listener. Gestures tend to precede the words that they depict; signaling and depicting less predictable words may help listener processing, by helping them get ready for and more easily recognize an otherwise hard to predict speech element (Holler & Levinson, 2019).

In fact, there is work showing that semantic gestures do help listener prediction and processing (Holle & Gunter, 2007; Zhang et al., 2021). But, the fact that gestures are functional for the listener, does not entail that speakers produce them for the listener. In principle, it is possible that speakers produce them when they themselves encounter a processing difficulty, potentially to help their own processing – and this may also, incidentally, help the listener. Here, we showed, for the first time, that gesturing to depict low probability words is not motivated by speaker needs: Speakers depict hard to predict words *for* the listener, instead.

4.4.3 Why do people use gestures to depict even when their gestures are not visible?

People use gestures to depict even when listeners cannot see them. This resilience of gesturing has been taken as evidence that semantic gestures are produced to help the speaker (Krauss & Hadar, 1999). However, we showed that the gestures speakers produce when they *cannot* be seen by a listener were not the same as the ones they produce when they *can* be seen – gestures did not depict low probability words when listeners could not see them.

Given that gestures do not depict lower probability words when they cannot be seen, why do they depict the words they do? One possibility is that semantic gestures that cannot be seen by a listener may simply reflect overlearned habits. People often gesture when they speak; therefore when they speak they cannot suppress gesturing even when gestures cannot be seen by a listener (De Ruiter, 2000). Due to this overlearned habit, gestures might surface anyway – depicting either a random content word in the sentence or a concept that is particularly salient in the speakers' imagery.

CHAPTER 5

CONCLUSION

Across three studies, I showed that people do not gesture when speaking is difficult, because gestures grow out of speaking difficulties or because gestures help people resolve those speaking difficulties. Instead, people gesture when speaking is difficult, because they are communicatively motivated. I showed that there is no reliable evidence to believe that people gesture, in part, to help their own speech production, since preventing gestures does not seem to harm speaking. Also, I showed, for the first time, that gesturing when speaking is difficult is not merely a symptom of speaker needs: People gesture when speaking is difficult for the listener, instead. People were more likely to gesture when they were disfluent or when they were about to utter a difficult word, only when those gestures could be seen by a listener. Speakers gesture when speaking is difficult, because they provide a communicative signal to the listener about the process and content of speaking: Gesturing during disfluent speech comment on experiencing problems with presenting an utterance to the listener and gesturing before a low probability word signals and depicts hard to predict words for the listener.

This dissertation adds to the work of many others showing that people gesture, in part, because gestures communicate to the listener. Here, I introduced novel communicative motivations for gesturing, adding to the many ways in which gestures can express meaning and be an important part of human conversation.

This dissertation, also, motivates a revision to theories of gesture claiming that people gesture, in part, because gestures serve a cognitive function in the speaker's mind. Gesture researchers long believed that people gesture, in part, because gestures help people speak – one way in which gestures can serve a cognitive function in the speaker's mind. Here, I systematically re-evaluated the empirical results that have been considered as evidence for this idea. I showed that the empirical evidence for this influential proposal is either not reliable (Study 1) or can be explained by alternative, communicative, mechanisms (Studies 2 and 3). So, contrary to long-held beliefs, helping speech production is not one of the cognitive functions of gesturing. Theories of cognitive functions of gesture will need to look beyond the role of gesture in speech production.

REFERENCES

Agresti, A. (2003). *Categorical data analysis.* John Wiley & Sons.

Akhavan, N., Göksun, T., & Nozari, N. (2016). Disfluency production in speech and gesture. In *Cogsci.*

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial cognition and computation*, *5*(4), 307–331.

Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of educational psychology*, *89*(1), 183.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*(2), 169–188.

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

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological science*, *22*(9), 1138–1144.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*(1), 31–56.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390–412.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, *4*, 328.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and social psychology bulletin*, *21*(4), 394–405.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse processes*, *15*(4), 469–489.

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

Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British journal of psychology*, *90*(1), 35–56.

Beattie, G., & Shovelton, H. (2000). Iconic hand gestures and the predictability of words in context in spontaneous speech. *British Journal of Psychology*, *91*(4), 473–491.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In *Recent advances in the psychology of language* (pp. 347–360). Springer.

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

Casasanto, D. (2008). Similarity and proximity: When does close in space mean close in mind? *Memory & Cognition*, *36*(6), 1047–1056.

Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right-and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351.

Casasanto, D. (2013). Gesture and language processing. *Encyclopedia of the Mind*, 372–374.

Casasanto, D., & Bottini, R. (2014). Spatial language and abstract concepts. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(2), 139–149.

Cienki, A. (2005). Image schemas and gesture. *From perception to meaning: Image schemas in cognitive linguistics*, *29*, 421–442.

Clark, H. H. (1973a). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior*, *12*(4), 335–359.

Clark, H. H. (1973b). Space, time, semantics, and the child. In *Cognitive development and acquisition of language* (pp. 27–63). Elsevier.

Clark, H. H. (1996). *Using language.* Cambridge university press.

Cook, S. W., Jaeger, T. F., & Tanenhaus, M. (2009). Producing less preferred structures: More gestures, less fluency. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 31).

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047–1058.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of memory and language*, *63*(4), 465–475.

Cravotta, A., Busà, M. G., & Prieto, P. (2018). Restraining and encouraging the use of hand gestures: Effects on speech. In *Proceedings 9th international conference on speech prosody* (pp. 206–210).

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of experimental psychology: General*, *122*(3), 371.

de Ruiter, J.-P. (2000). *The production of gesture and speech in mcneill d.(ed.), language and gesture (pp. 284–311).* Cambridge, England: Cambridge University Press.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, *23*(2), 283.

Finlayson, S., Forrest, V., Lickley, R., & Beck, J. M. (2003). Effects of the restriction of hand gestures on disfluency. *Proceedings of Diss, Gothenburg Papers in Theoretical Linguistics*, 21–24.

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30).

Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American journal of psychology*, *111*(1), 43.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in cognitive sciences*, *3*(11), 419–429.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological science*, *20*(3), 267–272.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological science*, *12*(6), 516–522.

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

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1-2), 147–172.

Hoetjes, M., Krahmer, E., & Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, *57*, 257–267.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: Erp evidence. *Journal of cognitive neuroscience*, *19*(7), 1175–1192.

Holler, J., & Bavelas, J. (2017). Multi-modal communication of common ground. *Why gesture*, 213–240.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior*, *33*(2), 73–88.

Holler, J., Turner, K., & Varcianna, T. (2013). It's on the tip of my fingers: Co-speech gestures during lexical retrieval in different social contexts. *Language and Cognitive Processes*, *28*(10), 1509–1518.

Hostetter, A. B. (2011). When do gestures communicate? a meta-analysis. *Psychological bulletin*, *137*(2), 297–315.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review*, *15*(3), 495–514.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). Does sitting on your hands make you bite your tongue? the effects of gesture prohibition on speech during motor descriptions. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 29).

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, *396*(6708), 228–228.

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

Kısa, Y. D., Goldin-Meadow, S., & Casasanto, D. (2021). Do gestures really facilitate speech production? *Journal of Experimental Psychology: General*.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? the gesture-for-conceptualization hypothesis. *Psychological review*, *124*(3), 245–266.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current directions in psychological science*, *7*(2), 54–54.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. *Gesture, speech, and sign*, 93–116.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of personality and social psychology*, *61*(5), 743.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive science*, *4*(2), 195–208.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1), 3.

Matthews, J. L., & Matlock, T. (2011). Understanding the link between spatial distance and social distance. *Social Psychology*, *42*(3).

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

McNeill, D. (2008). *Gesture and thought.* University of Chicago press.

McNeill, D. (2016). *Why we gesture: The surprising role of hand movements in communication.* Cambridge University Press.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 615.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American journal of psychology*, *117*(3), 411–424.

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for r. *Journal of Statistical Software*, *43*, 1–14.

Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech. *Language and gesture*, *2*, 186.

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological science*, *25*(4), 903–910.

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*(6), 747–754.

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, *34*(4), 602–619.

Ragsdale, J. D., & Fry Silvia, C. (1982). Distribution of kinesic hesitation phenomena in spontaneous speech. *Language and Speech*, *25*(2), 185–190.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological science*, *7*(4), 226–231.

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

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the smarc effect. *Cognition*, *99*(2), 113–129.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *Journal of personality and social psychology*, *89*(1), 1.

Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell system technical journal*, *30*(1), 50–64.

Team, R. C. (2020). R: A language and environment for statistical computing.

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

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of memory and language*, *50*(4), 395–407.

Wakefield, E. M., Hall, C., James, K. H., & Goldin-Meadow, S. (2018). Gesture for generalization: Gesture facilitates flexible learning of words for actions on objects. *Developmental science*, *21*(5), e12656.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American journal of psychology*, *114*(4), 591.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: A professional framework for multimodality research. In *5th international conference on language resources and evaluation (lrec 2006)* (pp. 1556–1559).

Yap, D., Brookshire, G., & Casasanto, D. (2018). Beat gestures encode spatial semantics. In *Proceedings of the 40th annual meeting of the cognitive science society.*

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Royal Society B*, *288*(1955), 20210500.