Listening Between the Lines: The Roles of Social Reasoning and Disfluency in the Comprehension of Ambiguous Utterances

Wei Li

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy
to
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
2024

Acknowledgements

The first time I visited Edinburgh, as a tourist, was in September 2015 – I still remember the moment when a strong blast of wind mixed with heavy rains hit my face. I struggled walking up the hills, while telling my friend that I would never come back to this city again. Who could ever imagine, three years later, I found myself studying here and spent the best of my six years. Despite the heavy wind and rain that still hit my face in winter, this city has become one of my favourite places in the world. All the wonderful people I have met here are the most important reason. Six years of this journey, where should I even start?

I would like to first express my deepest gratitude to my supervisors Martin and Hannah, who initiate my interest in psycholinguistics, and introduce me to disfluency and pragmatic research. They are the best supervisors one can ever imagine. Without their guidance, encouragement and support over the years, I would not be in my current position. They always provide the best suggestions, help me with everything they can, motivate me to step out of comfort zone, and encourage me to be more confident. Their endless great research ideas, patience with every student, and their modesty have always inspired me. Whenever thinking about them, I remind myself that they are the kind of academics I aspire to be some day. They have been the most precious part of this journey to me.

I am also grateful to every wonderful people in PPLS who have offered me generous help over the years. Thanks to Patrick and Dan for inspiring discussions about research ideas. Special thanks to Josiah and Umberto for always being incredibly helpful with statistical suggestions (And I really enjoyed and miss the days of tutoring with both of you!). Huge thanks to Alisdair for helping me debug my codes for mouse-tracking and eye-tracking experiments, and for being such a kind and patient Javascript teacher. Thanks to Esperanza for patiently helping me with eye-tracking when I first started my PhD and for always providing me with the most organised codes and scripts.

I also gratefully acknowledged the financial support from the PPLS research grant, which funded the experimental work presented in this thesis and enabled my attendance at conferences, such as XPRAG (2022, 2023) and AMLaP (2022, 2023).

On a personal note, I would like to extend my thanks to my parents, who have been supporting me and my every decision all these years. Without their strong support, I would not have had the courage to leave home to study in so many different places, from home to Beijing, and fly across the world to Sydney, now to Edinburgh. Thanks for always believing in me.

To the dearest friends I met in the UK – Ming, Janell, Elva, Zhuoni and Sugar – thanks for always encouraging me and being the best companion. To me, the time with you is the good weather in this city.

To all my best friends back home – Ren, Hao (M), Hao (F), Yinan, Yuan and Sishuo – thanks for always being there for me, even though we are far apart around the world. Thank you all for making space for me in your lives. A special dedication to Yuan, who always trusts me and provided enormous help when I first started learning statistics; and to Sishuo, who always encourages me and has been the best companion in pursing the path we insist.

To my best friend and partner, Hyunwoo. Thanks for always being at my side and becoming the warmest light in this foreign country. I am so glad we have made this journey wonderful together.

Finally, to two cherished people who have become stars, yet with whom I really wish I could share this moment:

To my grandpa: Grandpa, I hope you can see this, and I will always strive to be the granddaughter you can be proud of.

To Marian, my favourite lecturer at BFSU and a dear friend: Thank you for everything you have done for us and for your unwavering support. I will always aim to be the kind of academic and person that reflects even a little of your greatness.

It is indeed hard to condense six years' worth of gratitude into simple words, but all the wonderful people I have met during this journey have made the heavy winds and rain much more bearable, allowing me to call this city my second home. Thank you all for your support throughout this journey.

Abstract

Everyday speech is rarely fluent and often contains disfluencies, such as filled pauses *uh* and *um*. Disfluency is typically related to a speaker's production difficulties and it may therefore be regarded as an impediment to effective communication. However, disfluency can actually be a source of information, whereby a speaker conveys non-literal information, or paralinguistic cues (e.g., regarding the speaker's certainty or veracity), that can facilitate listeners' language comprehension in a given context. But exactly how does a speaker's manner of delivery exert its pragmatic effects on listeners' language processing?

This thesis takes a perspective to the study of disfluency by focusing on the influence of the speaker's manner of delivery on listeners' pragmatic comprehension, specifically the influence of context and disfluency on listeners' real-time interpretations of ambiguous statements. Four mouse-tracking experiments and one eye-tracking experiments are reported here, reflecting an innovative attempt at investigating the pragmatic roles of disfluencies on online language comprehension.

The experiments focus on the effects of filled-pause disfluencies like *uh* on listeners' interpretations of utterances. The mouse-tracking studies aimed to test how this type of disfluency influences listeners' interpretations of the ambiguous scalar quantifier *some*. The interpretation of *some* is known to depend on context, which makes this ambiguous word a good test case for exploring how pragmatic cues informed by disfluencies shape listeners' understanding. For example, how do listeners interpret statements like "I got, *uh, some* 'A's in my Psychology courses" in an interview context, where a speaker might be trying to present their few number of 'A's more positively, or deceptively?

The results – both in participants' final interpretations as well as in their online mouse trajectories – show that disfluency influences listeners' understanding of some

via a rapid integration of the social situation and the speaker's goals, and that this effect emerges in the early stages of listeners' comprehension. These results suggest that listeners engage in speaker-modelling behaviour rather than relying on a heuristic relation between disfluency and deception. Also, the current findings clarify prior work whose findings were compatible with either a social reasoning account or an account in which disfluency merely heightens listener attention. Further experiments explore whether listeners' interpretations reflect beliefs that the speaker has either stretched or even overwritten the meaning of *some*. Results from these experiments lead this work to broader questions about speakers' strategic use of language. Given the role of social reasoning, these findings open up new questions about possible differences in this kind of reasoning across individuals with varying pragmatic abilities or across contexts that evoke different conversational goals or speaker agendas.

A final eye-tracking study was conducted to test whether listeners' pragmatic reasoning varies systematically across individuals. The eye-tracking study was conducted with participants whose pragmatic abilities were posited to vary. The study elaborated on prior work that has suggested that similar behavioural outcomes might be achieved via distinct cognitive processes. People's eye movements were recorded as they responded to potentially misleading instructions to click on one of two objects which might conceal a hidden treasure. Eye movements analysis showed that listeners were less likely to click on the named object following disfluent instructions. Critically, when hearing disfluent utterances, a tendency to make early fixations to the named object increased with individuals' Autism Quotient scores. This finding provides further evidence that even where utterances are equivalently understood, the process by which interpretations are achieved varies across individuals with different pragmatic abilities.

Taken together, the results from mouse-tracking and eye-tracking studies reported in this thesis support an account of disfluency processing whereby listeners rapidly integrate the pragmatic cues of disfluency into a social reasoning process in a given context, a process that in turn varies with individuals' pragmatic abilities.

Lay Summary

In everyday conversation, people often use disfluencies like *uh* and *um* while speaking. Although these speech interruptions might seem like communication barriers, disfluencies can actually provide useful information about the speaker, such as their confidence or truthfulness. This thesis investigates how such disfluencies affect how listeners understand ambiguous statements – sentences that can have multiple interpretations, as the conversation is happening.

Five experiments were conducted: four using mouse-tracking and one using eye-tracking techniques. The mouse-tracking experiments aimed to understand how the disfluency *uh* influences listeners' interpretations, particularly focusing on the word *some*, which can have different meanings depending on the context. For example, in a job interview, a speaker says, "I got, *uh, some* 'A's in my psychology courses", listeners might interpret it differently based on the speaker's hesitations and the social context. Our results showed that listeners reason socially to understand the ambiguous utterances: they quickly integrate the presence of disfluencies with the social situation and the speaker's goals to understand the utterance, rather than just assuming from disfluencies that the speaker is being deceptive.

Additionally, an eye-tracking study explored whether people with different pragmatic abilities, as suggested by Autism Quotient (AQ) scores, interpret disfluencies differently. Previous research has suggested that people with higher AQ scores have difficulty accessing figurative meanings and tend to prefer literal interpretations of language, meaning that they often understand words and sentences exactly as the utterances are said. Pragmatic ability, reflected in lower AQ scores, involves understanding language in context, where words may carry implied or non-literal meanings. By exploring the relationship between AQ scores and how people comprehend disfluencies in utterances, we aim to better understand individual

differences in language comprehension. Our results showed that when hearing disfluent sentences, people with higher AQ scores tended to quickly lean towards a literal interpretation of what is being said. This finding suggests that the way interpretations are processed varies among individuals.

Overall, our findings highlight that listeners use disfluencies as pragmatic cues and they incorporate them into a social reasoning process - a process that varies with individual differences in pragmatic abilities.

Contents

List of Tables

List of Figures

Chapter 1
Introduction

When comprehending daily utterances, listeners not only take *what* is said, but also *how* the speech is uttered, into consideration. As a typical characteristic of spontaneous speech, disfluency is defined as any interruption in the flow of speech, including phenomena like hesitations, repetitions, corrections and filled pauses (e.g., *uh* or *um*). Disfluency is indicative of a speaker's manner of speech and is known to influence how listeners perceive an utterance (Fox Tree, 1995; Zuckerman, Depaulo, & Rosenthal, 1981). With a specific focus on filled pause type of disfluency *uh*, the effect of disfluencies on language comprehension is the topic of this thesis.

This thesis specifically examines the effects on comprehension of the disfluency *uh* in real-time language comprehension. Combining mouse-tracking and eye-tracking methods, the studies here explore in depth how disfluencies affect online pragmatic comprehension, testing the combined effects of manner of delivery and social context. By analysing not only the comprehension result, but also the process of comprehending, the current results emphasise the point that even when the utterances are understood equivalently, the process by which an interpretation is achieved can vary.

Other than the widely used eye-tracking method, mouse-tracking paradigms are applied in the current study of online language processing. Despite its potential, mouse-tracking has not yet been extensively adopted in language comprehension research. Like the visual world paradigm, mouse-tracking is able to capture the dynamics of language processing in real time (Spivey, Grosjean, & Knoblich, 2005). Deviations and hesitations in mouse movements may hint at features of language processing, for example, how comprehenders react to ambiguity or other conflicts of interpretation. These characteristics make this technique a perfect method in this thesis to examine the real-time effect of disfluency in language comprehension.

Building upon previous research, this thesis presents a series of experiments designed to investigate the pragmatic effects of disfluencies in interpreting ambiguous statements. These experiments focused on a specific ambiguous word – the scalar quantifier *some*. Listeners' interpretations of the word *some* depend largely on context (e.g., Bonnefon, Dahl, & Holtgraves, 2015; Breheny, Katsos, & Williams, 2006; Politzer-Ahles & Husband, 2018). Therefore, studying how *some* is interpreted in a given social context provides an appropriate testbed for how disfluency exerts pragmatic cues for listeners' comprehension. This series of studies indicate an account of how listeners comprehend disfluencies: disfluency cues a rapid social reasoning process where listeners integrate the pragmatic cues of disfluency fast in a given social context. This process includes two stages of reasoning that listeners reach an initial interpretation compatible with the broad semantics of the utterance, but can then override this if context and disfluency allow.

With the finding from mouse-tracking studies that disfluency cues a social reasoning process, a further question arises as to whether this reasoning process varies across individuals. An eye-tracking study explored this question and tested the processing differences among individuals with varying levels of pragmatic ability, suggested by Autism Quotient scores. This study provided further evidence that disfluency comprehension cues a social reasoning process rather than an automatic process as it varies with individual's social cognitive profile.

1.1 Thesis overview

Chapter 2-4 provide the background for the experimental work presented in this thesis. In Chapter 2, I review the disfluency literature. I introduce relevant concepts, define disfluencies in the context of this thesis, and describe the possible causes of disfluencies. The focus of this chapter is on how disfluency in spontaneous speech affects language comprehension, where I discuss and evaluate existing evidence and different accounts regarding the disfluency comprehension process. In Chapter 3, I then review literature relevant to the comprehension of utterances with ambiguity, with a focus on the ambiguous word *some*. I discuss and evaluate various accounts of the processing of *some* and explain how the effect of disfluency can be explored via examining the processing of this word. Chapter 4 introduces mouse-tracking and eye-tracking methods. Focusing on the main methodology used in the experimental work, I explain how mouse-tracking works and why the mouse-tracking data and methodology is suitable for investigating language comprehension process. By reviewing relevant literature on language comprehension that has applied mouse-tracking methods, I evaluate a commonly used mouse-tracking paradigm and techniques for analysing mouse-tracking data.

Chapter 5-7 comprise the experimental studies and are the main body of this thesis. Chapter 5 and Chapter 6 contain a series of mouse-tracking studies, investigating the effects of the filler pause disfluency *uh* on listeners' interpretations of *some*. In Chapter 5, I describe two mouse-tracking experiments which explored whether the pragmatic effects of disfluency on the processing of *some* observed in previous literature persist across different social contexts. The results provide additional evidence for the pragmatic role of disfluency in real-time language processing. The findings also highlight that disfluency influences listeners' processing of *some* via modelling the speaker's mind to reason socially for appropriate interpretations in the given context, rather than heuristically. In Chapter 6, I describe two more mouse-tracking experiments which further investigated listeners' ability to engage in social reasoning to model their interlocutor. The results affirm a social reasoning account regarding the effect of disfluency on comprehension, indicate the speaker's strategic use of language. In Chapter 7 I present an eye-tracking experiment. Results from previous mouse-tracking experiments raise the possibility of listeners' pragmatic ability influencing the effect of disfluency on comprehension. In this eye-tracking study, I test the hypothesis of individual differences in pragmatic reasoning process. Eye-tracking results show differences of comprehension process vary by individuals' pragmatic abilities. Similar comprehension outcomes are shown to arise via distinct cognitive processes.

In Chapter 8 I summarise the main findings from the experimental work. I suggest what functions of disfluencies may have driven the observed results in language processing and describe a possible account of disfluency comprehension to explain the findings, incorporating the listeners' effective use of social cues and their own pragmatic abilities in comprehension.

Chapter 2
Disfluencies in Speech

2.1 Introduction

Spontaneous speech in daily conversations, which is the prototypical form of human language use, is rarely fluent. Rather than the fluent speech studied to build language production and comprehension models, disfluency is the actual norm of daily language and a pervasive phenomenon that has gradually attracted psycholinguists' attention.

Disfluencies are often caused by difficulties in the speaker's production process. As for the listener who hears someone produce disfluent speech, they may on one hand attribute that disfluency to production difficulties, but they may also draw other conclusions about what gave rise to the disfluency. Moreover, this effect of disfluency on listeners' language comprehension seems to vary by contexts. This thesis has two broad foci: (i) the question of how listeners use disfluencies as pragmatic cues to reason socially and (ii) at what stage of comprehension this reasoning happens.

To understand what previous studies have covered for these points, this chapter reviews the disfluency literature with the aim of providing relevant background for the experimental work presented in this thesis. I define disfluency in the context of this thesis, briefly introduce various types of disfluencies with a focus on filler disfluencies. I explain the possible causes of disfluencies and discuss the communicative role disfluencies play in a speaker's production. I discuss findings from previous studies exploring the question of how disfluency influences listeners' comprehension process. I highlight two accounts to address this question and emphasise the beneficial effect of disfluencies on acting as pragmatic cues which listeners are able to integrate quickly in comprehension process.

2.2 What are disfluencies?

2.2.1 Define disfluencies

Utterances have been described as being produced in a three-stage pipeline[1]: conceptualisation, formulation and articulation (Levelt, 1983, 1989). Before uttering a sentence, speakers must first conceptualise the message they want to convey. Formulating this message into speech involves planning the syntactic structure, retrieving appropriate lexical words, properly inflecting the words with correct grammatical forms, and detailing the articulation and phonetics. The phonological form resulting from this process can eventually be articulated. A self-monitoring process also accompanies these three stages. Speakers monitor their internal speech before articulation and also monitor their external speech after articulation.

Moreover, language production is generally recognised to be an incremental process, with speakers conceptualising and formulating the conveyed messages simultaneously when speaking (e.g., Ferreira & Swets, 2002; Marslen-Wilson & Tyler, 1981). Therefore, it is expected that difficulties may arise during any stage of production. These difficulties in production stages can eventually lead to disfluencies.

Disfluency is defined as a phenomenon characterised by interruptions in the flow of spontaneous speech that do not add propositional content to utterances (Fox Tree, 1995). Disfluencies are a pervasive feature of daily communication, occurring with an estimated rate of six in every hundred words excluding silent pauses (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995; Shriberg, 1994). Even though fluency is typically the speaker's goal, disfluency is the actual norm, leading to the rising importance of exploring the effects of disfluencies on listener's language comprehension.

Dependent on the field of study, the use of the term *disfluency* itself varies. For example, disfluency (also as dysfluency) in speech pathology specifically refers to disorders of fluency, such as stuttering (see Lickley, 2015 for detailed definitions); alternatively, disfluency is regarded as a lack of proficiency in a target language when used in second language learning studies (De Jong, 2018; Kormos & Dénes, 2004).

[1] For the convenience of explaining, this thesis assumes the basic three-stage speech production model raised by Levelt (1983, 1989).

The review in this thesis of the disfluency literature is not intended to be an exhaustive review, but to define disfluencies and related terms used throughout this thesis with a focus specifically on disfluencies produced in typical, healthy adult native English speakers, given that the aim of the thesis is to explore how these disfluencies are perceived by listeners. Common types of disfluency phenomena produced in daily speech are introduced in detail in the next section.

2.2.2 Types of Disfluencies

There has been little consensus on how to categorise disfluencies in spontaneous speech. Defined as breaks involved in the flow of speech, disfluencies are further categorised into two main phenomena resulting from problems in the production process: (i) disfluencies that simply introduce *hesitations* in production; or (ii) those that alter or *repair* part of the utterance. Each category includes subtypes which are detailed in Table 2.1. This typology adopts an approach set out by Lickley (2017) and Shriberg (2001).

Table 2.1 Two main categories of disfluencies with detailed subtypes and examples (Italics indicates the interruption point)

Disfluency category	Disfluency type	Example
Hesitations	Fillers (Filled pauses)	I ate, *uh*, some oreos.
	Silent pauses	Do you truly *(silence)* believe so?
	Lexical fillers	She was, *you know*, very nice.
	Prolongation	I have read *proooobably* more than this.
	Repetitions	That was *all*, *all* I can say.
Repairs	False starts	It is *[summ], autumn* now!
	Substitutions	Do you have *any health cover, any health insurance*?
	Insertions	*I agree, I also agree* with you.

The subtypes listed in Table 2.1 refer to the type of wording change between the lexical content before and after the interruption point, without attempting to identify the causes or functions of these disfluencies, but rather merely showing the differences between the observable patterns of these disfluency types in interrupting the flow of utterances. As the experimental work presented in the current thesis focuses on the category of fillers (i.e., *um* or *uh*), a general introduction to this specific kind of disfluency is presented below. Further details of filler production and comprehension are discussed in the later sections.

2.2.3 Fillers

Fillers, also referred to as filled pauses, are a type of hesitation disfluency characterised by a delay of utterance production followed by continuing but not repairing the utterance (Macgregor, 2008). The phonological form of fillers varies across languages[2] and accents. In English, fillers are commonly represented in the orthographical form of *um* or *uh*, with an American English pronunciation of *um* or *uh* and a British English pronunciation *er*. For the convenience of discussion, I make no distinction among these variants of fillers, and I refer to fillers using the term specified by the authors in the relevant literature. When discussing the current studies throughout this thesis, I use *uh*[3] specifically to represent the filler.

Fillers were once taken as mere performance errors, simple production problems that contain no linguistic meaning (Blacfkmer & Mitton, 1991; Chomsky, 1965; Cossavella & Cevasco, 2020). However, despite a representation as a speech error, accounts arose suggesting that the production of fillers can provide a window into underlying speech production mechanisms (Goldman-Eisler, 1961). For example, the location of fillers can reflect speakers' decision-making process in at least lexical and structural levels (e.g., Rochester, 1973). Also, fillers are regarded as non-linguistic signals used to hold the floor between conversation turns (Maclay & Ossgood, 1959). Apart from these accounts regarding fillers as non-lexical terms, Clark and Fox Tree (2002) proposed that *um* and *uh* are conventional English words, as they found from analysing large corpora of spontaneous speech that speakers monitor their speech plans for delays, formulate and produce fillers as for any words. Under this account, fillers no longer just signal a production problem, but they carry communicative goals and meanings. This viewpoint has not gone unchallenged (see Corley & Stewart,

[2] Articles discussing fillers in different languages include the following: French: Crible, Degand, & Gilquin (2017); Dutch/German: De Leeuw (2007), de Boer & Heeren (2020); Portuguese: Moniz (2013); Chinese: Yuan, Xu, Lai, & Liberman (2016); Japanese: Watanabe, Hirose, Den, & Minematsu (2008).
[3] Many previous studies have also taken the view that the use of *um* and *uh* in the spontaneous utterances should be distinguished (e.g., Clark & Tree, 2002; Fox Tree, 2001), but this thesis does not explore this perspective in detail and uses *uh* as a general representation of fillers.

2008), but the communicative role of fillers provides us an important reason to explore them in spontaneous speech.

2.3 Why do speakers produce disfluencies?

As discussed in previous section, this thesis focuses on the perception of the disfluent utterances. By emphasising listeners' perspectives, I highlight the importance of exploring how disfluency is understood. This does not mean a complete neglect of the speaker's side. Before moving on to the discussion of the comprehension of disfluencies, it is necessary to discuss first why a speaker might produce disfluencies (especially fillers) to understand betters how a listener perceives these disfluencies in processing.

Disfluencies often signal that a speaker is having difficulties in speech production and such difficulties can occur in different stages (i.e., planning, lexical retrieval or articulation) (Schnadt & Corley, 2006). It has been proposed that different types of disfluencies may signal different issues during production (Bortfield et al., 2001). For example, repairs stand out from other disfluencies as they correct a speech error rather than serve as a mere interruption.

Moreover, the production of disfluencies varies among individuals. Speakers with different speech speeds may differ in their production of disfluencies: individuals who speak slowly are more likely to use fillers, silent pauses or repetitions, while those who speak faster tend to produce more repairs (Ferreira, Lau, & Bailey, 2004). Individual differences in executive functions and intelligence also lead to differences in the types of fillers produced. For example, inhibitory control is related to repair disfluencies production (Engelhardt, Nigg, Ferreira, 2013), memory abilities are linked to production of silent pauses and repairs, while verbal intelligence is found to be related to repetitions (Engelhardt, McMullon, & Corley, 2019).

With a focus on filler production, I summarise the evidence generally describing why speakers produce disfluencies in the sections below. I also expand on the observation that speakers produce disfluencies as communicative signals.

2.3.1 Causes of filler production: Difficulties

2.3.1.1 Of speech planning process

When planning difficulties arise, a speaker may produce fillers, particularly at the beginning of an utterance or a phrase. During the conceptualisation and formulation process in speech production, choices of utterances, attentional resources allocated to formulate speech, as well as the ability to self-monitor, are all important factors that may be related to filler production.

Choices a speaker faces during planning can lead to disfluency production. Beattie and Butterworth (1979) found more fillers before words with low contextual probability, and they considered this to be because when speakers choose words that are less predictable or obvious in a given context, they might be more conscious of the word selection process. This heightened awareness in deciding which word to use leads to more disfluencies.

The point that the increase of fillers rate is related to ambiguity due to a wider range of choices has also been supported in Christenfeld (1994). In this study, participants were asked to describe a path through a maze so that another person, upon hearing the description, could accurately recreate it. Each participant was provided with three mazes, each containing varying numbers of possible route choices. The results showed that filler production was related to the number of choices: speakers produced a higher rate of filled pauses when describing the maze with more alternative routes.

Schachter et al. (1991, 1994) also illustrated how the choices of words were involved in disfluency production. Schachter et al. (1991) found that the frequency of filler usage varied with academic disciplines being lectured on, where lecturers in the natural sciences were observed to use fewer fillers compared to those in the humanities, with social sciences falling in between. Schachter et al. hypothesized that this variation in filler occurrences might be due to the linguistic choices available in each field – specialised and limited vocabulary in natural sciences may reduce the cognitive load on lecturers when choosing appropriate terms. Schachter et al. (1994) analysed vocabulary size across different types of texts related to these fields, which further supported the point that natural sciences tend to use a narrower range of terms.

Availability of attentional resources are also related to filler production. In Oomen and Postma (2001), they used a dual-task paradigm to manipulate whether a speaker's attention is divided or not. Speakers who performed a picture story-telling task produced more fillers when they simultaneously performed a tactile-form recognition task. The concurrent tasks were speculated to limit attentional resources for speech formulation and the increased number of fillers was regarded as an automatic reaction to the increased planning difficulties when encountering concurrent tasks.

The production of fillers has also been related to successful self-monitoring, whereby a speaker detects and corrects mistakes made during planning or articulation (Postma, Kolk, & Povel, 1990; Postma & Kolk, 1992; Postma & Noordanus, 1990). In production studies conducted by Postma et al. (1990) and Postma and Noordanus (1990), speakers were instructed to focus on their speech, which was expected to boost self-monitoring. Both studies found that speakers made fewer speech errors while maintaining the same numbers of fillers, resulting in an increased filler to error ratio. Conversely, Postma and Kolk (1992) found when reducing the self-monitoring process by masking auditory feedback with noise, speakers produced fewer fillers. These findings suggest that when speakers increase self-monitoring, they avoid speech errors at the cost of producing more fillers.

2.3.1.2 Of speech content

Besides their association with planning difficulties, the production of fillers has also been found to relate to certain types of messages being conveyed. In this sense, the occurrences of fillers may reflect information beyond speech itself, such as speaker knowledge.

The presence of fillers is associated with reduced confidence on the part of the speaker. Smith and Clark (1993) showed that speakers were more likely to give incorrect answers to general knowledge questions when their responses included fillers compared to when their responses were fluent. Furthermore, using Feeling of Knowing (FOK) measures (Hart, 1965), they found that speakers rated themselves as less confident about their answers when they had been disfluent.

Moreover, previous studies have also found that speakers produce more fillers before content words (Maclay & Osgood, 1959), low-frequency words (Beattie, 1979; Levelt, 1983), less-preferred syntactic structures (Cook, Jaeger, & Tanenhaus, 2009), and discourse-new or unfamiliar expressions (Arnold, Losongco, Wasow, & Ginstrom, 2000; Arnold, Kam, & Tanenhaus, 2007).

The relationship between filler production and the speaker's conveyed message not only explains some occurrences of fillers, but also provides a strong basis to explore the informative role of fillers in spontaneous speech. This role of fillers is discussed below and further in section 2.4.

2.3.2 Fillers' communicative role

As mentioned in introducing fillers in section 2.2.3, it has been controversial whether fillers have a functional role in communication rather than merely being audible errors due to speech planning or production difficulties. Goodwin (1980) notes that, while disfluencies in speech are typically seen as signs of difficulty in formulating speech on the part of the speaker, they show the speaker's competence in constructing sentences that are sensitive to the listener's needs and reactions. For example, fillers can be used by the speaker as a device to hold the floor in dialogue, informing listeners not to interrupt and that the speech is expected to continue. In this way, fillers are not simply a by-product of difficulties in speech production but can also be the intended output when communicating information about the status of the discourse to the addressee.

Fillers can also be used by a speaker to signal upcoming delays. With a detailed corpus analysis of natural language, Clark and Fox Tree (2002) demonstrated that fillers, *um* and *uh*, were systematically used to indicate lengths of upcoming delays, with *um* used to signal longer pause in upcoming speech and *uh* signalling a shorter upcoming pause. Speakers are also found to use fillers strategically to maintain a continuous speech stream and avoid a silent pause. Jefferson (1974) found that speakers use the filler *uh* when they anticipate potential error or difficulty in their upcoming speech.

Speakers' modulation of their filler use according to their speech partner also provides evidence for fillers' communicative intent (Oviatt, 1995; Walker, Risko, & Kingstone, 2014). Oviatt (1995) examined participants' spoken disfluency rates in both human-computer and in human-human interactions. Filler rates were found to be significantly lower when participants spoke to a computer compared to speaking to another human, whether via telephone conversations or in face-to-face dialogues.

Walker et al. (2014) also investigated this phenomenon by asking participants to answer a series of general knowledge questions posed by either a human (human-human interaction) or a computer (human-computer interaction). They found that participants produced more fillers when responding to a human. Additionally, Walker et al. introduced a condition where another human was present during human-computer interactions. They discovered that filler rates were higher when another human was present compared to when participants were alone while answering questions posed by computers. These differences in filler rates can be explained by various factors. For example, social factors like social pressures and increased anxiety when another human is present may lead to more filler production. Furthermore, the presence of another human partner may suggest an expectation for a timely response, and fillers can be used by speakers to indicate that they are actively working on a response, albeit with a slight delay.

Speakers were also found to control the use of fillers to make themselves more trustworthy for listeners (Arciuli, Mallard, & Villar, 2010; Loy, Rohde, & Corley, 2018). In a treasure-hunt game setting where speakers were asked to mislead listeners about the locations of treasures, speakers were found to produce fewer fillers when giving false statements (Loy et al., 2018). Arciuli et al. (2010) also found a decrease in speakers' filler use when lying compared to truth-telling. These results suggest that when producing disfluencies, speakers take account of listeners' expectations and manipulate their own behaviours to control listeners' belief towards them.

The evidence from the studies above suggests that fillers can be a communicative signal, but this cannot be fillers' only role in the speech. They can still be an indication of task or planning difficulties. Moreover, although fillers were found to have higher production rate in dialogues, they do still appear in monologues which are not communicative situations (Oviatt, 1995; Finlayson & Corley, 2012). This can be explained by considering monologues as conversations with "a virtual partner" (Clark, 1999) in mind. However, the presence of fillers in non-dialogue situations suggest they may serve other functions. Despite other interesting possibilities, I would like to emphasise the communicative role for the topic of this thesis. Speakers do not need to intend to produce fillers, and fillers' occurrence can certainly convey information to listeners, which is discussed further in later sections regarding the comprehension of disfluencies.

2.3.3 Summary: the production of filler disfluencies

The occurrence of disfluency is commonly associated with planning difficulties in speech production. The production rate of disfluencies increases when speakers have choices to make, have less attention available, or when speakers are conscious of self-monitoring during production. Fillers, other than an indication of production difficulties, are also regarded as containing communicative information. Speakers use fillers (either intentionally or not) to hold the floor, and the speech following fillers conveys specific messages about the speaker's knowledge or the content of the speech. Bearing in mind why a speaker produces disfluencies, in the remaining sections, I consider fillers from a listener's perspective with a focus on how fillers exert influence on speech comprehension.

2.4 How do disfluencies affect listeners' language comprehension?

An intuition when it comes to listeners' perception of disfluencies in speech is that people rarely notice them and may even "filter" them (Martin & Strange, 1968) when interpreting speech. Alternatively, disfluencies may add a burden to listeners' grammatical and semantic understanding of the speech considering the definition of breaking the flow of perception. However, disfluencies can do more than provide listeners with information about the speaker's difficulties in production; they can also influence listeners' judgement of the speaker themselves or the listener's processing of speech, which can benefit, instead of hindering, listeners' comprehension process by providing communicative information.

For example, in Brennan and Williams (1995)'s question-and-answer tasks, the presence of fillers was found to be interpreted by listeners to estimate the speaker's knowledge. The more fillers a speaker produced in spontaneous verbal responses, the more likely the listeners were to rate that speaker as less knowledgeable and less confident about their answer. Brennan and Schober (2001) also showed that the information available from the presence of disfluencies can help listeners manage and adjust to interruptions that naturally occur in spontaneous speech. In a referential communication task, Brennan and Schober asked listeners to select a referent from a set of geometric objects based on instructions with fillers preceding the targets (e.g., "Move to the yel-*uh*, purple square") or without fillers (e.g., "Move to the yel, purple square"). They found that listeners were quicker and equally accurate in selecting the targe image when hearing mid-word interruptions accompanied by fillers (e.g., "Move to the yel-*uh*, purple square") compared to those without fillers. The effects of disfluencies on listeners' judgements about the speaker and the content of the speech suggest that listeners are sensitive to disfluencies and use the communicative information disfluencies provide. However, what specific information disfluencies convey is not entirely clear; they may provide cues affecting listeners' attentional processes or they may signal pragmatic information. It remains underexplored how listeners become aware of and use this information during comprehension. At what stage of processing does the disfluency exert its influence on listeners' comprehension – can the effect of disfluency be observed both in the *outcome* and the *process* of comprehension?

In the following sections, I review evidence from studies investigating the effects of disfluencies (with a focus on fillers) on speech comprehension. I first introduce two general accounts of how listeners possibly make use of the communicative information in disfluencies, and touch upon the time-course of listeners' comprehension. I then discuss the relationship between disfluency and social contexts, emphasising disfluency's role as a pragmatic cue within a listener's social reasoning process. Lastly, I discuss the pragmatic features of disfluency, whereby disfluency can provide a cue to potential speaker deception, or it may invite heightened attention on the listeners' part, or it may influence the interpretation of ambiguous utterances.

2.4.1 Two accounts: heuristic or perspective-taking?

Disfluency interpretation via heuristics

One way of understanding the impact of disfluency on listeners' interpretations is to assume a heuristic approach, where disfluencies inherently provide listeners with default inferences about the utterances. For example, utterances with fillers have been found to bias listeners' perceptions of speaker honesty, with disfluent speakers being rated to be less honest (Fox Tree, 2002). Additionally, listeners associate fillers with speaker confidence, perceiving speakers who produce disfluent utterances as less confident when answering questions (Brennan & Williams, 1995).

Such findings can be understood within an account in which disfluencies impose a default interpretation bias of heuristic. This heuristic leads listeners to an interpretation without requiring that they engage in a reasoning process. In the cases of disfluency being judged as a pragmatic cue to lying, previous studies support a heuristic view with the evidence that the effect of disfluencies on listeners' is quick (Zuckerman, Koestner, & Driver, 1981) and listeners' judgement of the causes of a speaker's disfluency is not always correct (DePaulo, Rosenthal, Rosenkrantz, & Green, 1982; Loy, Rohde, & Corley, 2018). Moreover, listeners appeared to make a heuristic judgement, as shown by DePaulo et al. (1982) whose study established a mismatch between a speaker's actual disfluency rate when lying and listeners' perception of deception from disfluent speech. While the number of filled pauses (*uh* or *um*) that speakers produced did not differ when they told the truth and when telling lies, listeners always judged the utterances with higher rates of filled pauses as a sign of lying.

However, the aforementioned findings can also be interpreted from another perspective whereby the listeners' inaccurate judgements in fact reflect a more nuanced model of the speakers. Listeners may think back to the situations when they themselves have been speaking, with the belief that disfluencies arose more often with cognitively effortful actions such as lying. This argument opens the possibility of a perspective-taking account of comprehension which may require listeners to consider speakers' metacognitive status and reason about that status may influence the production of disfluencies.

Disfluency interpretation via perspective-taking

In real-life conversations, we take turns playing the role of speaker and listener, and therefore speakers' production of disfluencies may provide a window for understanding how disfluencies are comprehended. Listeners, when they become the person comprehending the utterances, may attempt to review the situations when they are the speaker and are speaking disfluently. In this way, listeners take the perspective of speakers[4] to understand what the speaker's disfluencies may mean in a given context.

[4] See Brown-Schmidt (2009), Brown-Schmidt and Hanna (2011) for systematic reviews of how listeners combine knowledge of the speaker's perspective with contextual information in real-time processing to reach their ultimate interpretation.

It should be noted that listeners may be taking the perspective of a specific speaker (i.e., the actual individual who is communicating in that moment) or a generic speaker (i.e., an abstract, non-personalised speaker who could represent any typical communicator in a similar situation). Distinguishing between these two possibilities is not the focus of the current thesis. Therefore, we are using the term ‘perspective-taking’ loosely throughout the thesis, without implying a strong distinction between speaker-specific and speaker-general reasoning. Instead, the emphasis is placed on how interpretations are shaped by the broader social context and the desirable outcomes within that context, rather than the mental modelling of a specific individual speaker.

This perspective-taking account assumes that disfluencies signal a delay in the forthcoming speech and that listeners reason about this delay from the contextual information and their cooperative assumptions about dialogue (Clark & Fox Tree, 2002). According to this account, listeners’ comprehension of disfluent utterances should be speaker- and context-specific, depending on their understanding of a specific speaker and a specific social context.

Part of the evidence supporting the perspective-taking account comes from studies on individuals with Autism Spectrum Disorders (ASDs), who are known to have challenges with Theory of Mind (Baron-Cohen, 1995) and may thus struggle to understand what another person is thinking or feeling (e.g, Hallin, Garcia, & Reuterskiöld, 2016; Lake, Humphreys, & Cardy, 2011). Lake et al. (2011) analysed language samples from adults with ASD and found that individuals with ASD produced fewer fillers compared to individuals with typical development. Similarly, Hallin et al. (2016) found that school-age children with ASD used fewer fillers in conversations compared to their typically developing peers. The reduced use of fillers by individuals with ASD (whose perspective-taking abilities may differ from neurotypical individuals) could be taken to suggest that fillers serve a listener-oriented function, whereby they are produced for the benefit of listeners. For individuals with difficulties in perspective-taking, this lack of attention to social signals involved in fillers may result in challenges understanding the implicit social meanings conveyed by fillers (Irvine, Eigsti, & Fein, 2016). The differences between ASD and non-ASD individuals will play a role in this thesis in an experiment that tests possible cognitive resources and mechanisms that are implicated in a perspective-taking account (see Chapter 7).

A study by Arnold and colleagues (Arnold et al., 2007) also provides supporting evidence for the perspective-taking account. If a listener hears a speaker being disfluent (e.g., "click on thee uh…"), the listener may guess that the speaker has referred to a new or unusual object that is difficult to describe rather than an everyday object (such as an apple). However, if a listener is given speaker-specific information, such as that the speaker has object agnosia and thus naturally has difficulty naming everyday objects, the listener no longer anticipates the relationship between speaker's uttering disfluencies and referring to the unusual-shaped object. Results from Barr (2001), together with Barr and Seyfeddinipur (2010), also showed that listeners were more likely to expect an upcoming mention of a new referent when hearing disfluent speech with filler *um* than when hearing the same utterance where the filler was replaced by noise. This effect of disfluency was also found to be dependent on speaker, as it depends on what is regarded to be new for the speaker, rather than for the listener. These studies provide further evidence for a perspective-taking mechanism underlying the effect of disfluencies and indicate that listeners make inferences about the speaker's knowledge and mental state to interpret disfluent utterances.

These two accounts, perspective-taking and heuristic, offer different perspectives on how listeners utilise disfluencies in comprehending utterances in real time. A critical difference between these two accounts is whether context influences the effects of disfluencies. The heuristic account assumes that context does not affect listeners' comprehension of disfluent utterances. For example, when judging the truthfulness of utterances, disfluent utterances are perceived as deceptive regardless of contexts. In contrast, the perspective-taking account suggests that the effects of disfluency are context-specific, meaning that the pragmatic cues conveyed by disfluencies depend on the given context. In this account, for example, listeners do not automatically judge disfluent utterances as untruthful but interpret them through a reasoning process. In the heuristic account, the fast integration of disfluency into interpretation does not necessarily eliminate all involvement of speaker modelling, but such speed could also be evidence that modelling the speaker is a fast process when appropriate context is provided. In the next section, I introduce the role of context in listeners' comprehension in detail.

2.4.2 Role of context

Context is a critical factor to discuss when considering language comprehension, as in daily conversations, listeners do not interpret the speaker's words in isolation, but they also consider the context in which those words are used (Garten, Kennedy, Sagae, & Dehghani, 2019). As suggested by the perspective-taking account, a speaker's manner of speech, (i.e., whether the speaker is fluent or not) can have different effects on speech comprehension in various social contexts. For example, fillers are often used strategically in speech to gain time and enhance overall communicative competence. Listeners are more likely to make negative evaluations towards speakers if fillers are produced in formal speaking contexts, such as job interviews and public speaking (Rose, 2008; Brosy, Bangerter, & Mayor, 2016).

This pragmatic function of disfluencies suggest that context may influence disfluency's effect on listeners' online comprehension via a *social reasoning process*. By social reasoning, I refer to a process where listeners are cued by the occurrences of disfluencies to pragmatically reason why the speaker is disfluent in a specific social context.

In the following sections, I elaborate the types of social cues that disfluency can provide to listeners. I summarise several kinds of social cues disfluencies can act as: pragmatic cues of deception; cues to heighten attention and cues for clarifying ambiguous utterances.

2.4.3 Disfluencies as pragmatic cues

2.4.3.1 Disfluency cues deception

Disfluencies, such as fillers, have been found to influence listeners' pragmatic inferences about whether a speaker is truthful or deceptive. Studies on lie detection draw attention to vocal cues when recognising deception (Bond & DePaulo, 2006) and in a questionnaire task used in Zuckerman and colleagues (Zuckerman, Koestner, & Driver, 1981), fillers were found to be one of the vocal cues that listeners frequently associate with perception of lying. In a meta-analysis of 33 studies assessing the cues listeners use to detect lies, the presence of fillers in speech was consistently identified across different studies as a significant indicator of deception (Zuckerman, DePaulo, & Rosenthal, 1981).

Recent studies have also provided evidence relating speakers' disfluency to listeners' pragmatic interpretation of deception in real-time language processing (Loy, Rohde, & Corley, 2017; Loy, Rohde, & Corley, 2018). Loy et al. (2017, 2018) used a treasure-hunting computer game task where listeners needed to click one of two objects, while listening to a potentially dishonest speaker giving instructions in fluent ("The treasure is behind the giraffe") or disfluent sentences (e.g., "The treasure is behind the, uh, giraffe"). Listeners' eye and mouse movements were found to show a bias towards the speaker's named image (referent) when hearing fluent utterances and towards the unnamed image (distractor) in disfluent condition.

Disfluencies cuing deception may not be the only factor affecting listeners' biases on distractor in previous studies. There is evidence that listeners are also sensitive to the context in which disfluencies occur, particularly when evaluating whether the speaker is lying or not (King, Loy, & Corley, 2017). King et al. (2017) used the treasure-hunt game paradigm of Loy et al. (2016, 2017), manipulating not only the manner of speech of the speaker, but also an alternative local cause of speaker disfluency – whether the speaker was distracted by a car horn. Listeners' eye and mouse movements towards the referent indicated that participants judged the speaker to be truthful, while movements towards the distractor indicated a judgement of lying. King et al. observed that listeners initially moved towards the referent and only fixated and clicked the distractor later when car horn is present with the speaker's disfluent instruction. This pattern showed that listeners initially judged speaker to be truthful but changed their judgement, concluding the speaker was lying. This result suggests that listeners might temporarily consider contextual information, such as car horn, and reason that the distraction caused the speaker's disfluency. This finding indicates that listeners' disfluency-deception bias is not applied as an across-the-board heuristic. While disfluency can act as a cue for deception, when different contexts are provided, listeners are capable of modelling the speaker's mind to reason about the cause of disfluency in that specific context.

2.4.3.2 Disfluency cues to heighten attention or to disambiguate

Besides cuing potential deception, disfluencies have also been found to have functions beneficial for listeners' comprehension in spontaneous speech, including heightening listeners' attention towards speech that follows (Collard, Corley, MacGregor, & Donaldson, 2008) and informing listeners of resolution of ambiguous utterances (Bailey & Ferreira, 2007; Loy, Rohde, & Corley, 2019).

Collard et al. (2008) utilized event-related potentials (ERPs) and asked participants to listen to speech with hesitations while these listeners' brain responses were recorded. They found that words preceded by a disfluency were more likely to be remembered and their results highlighted disfluency's relationship with attention in language processing that disfluencies can provide cues to heighten attention for words following disfluencies.

In a visual world paradigm task by Bailey and Ferreira (2007), participants were asked to manipulate the objects placed in front of them following ambiguous utterances with fillers placed in different positions, such as "Put the, uh uh, apple on the towel in the box" vs. "Put the apple on the, uh uh towel in the box". Listeners' eye movements and their reactions showed more fixations towards the towel in the box in the late disfluency condition compared to the early disfluency condition. This finding indicates that disfluencies can inform the resolution of referential ambiguities and therefore may act as cues to disambiguate syntactic structures.

Disfluencies' role as a cue to influence the interpretations of ambiguous utterances is also shown in Loy et al. (2019). Loy et al. explored how listeners' interpretations of the ambiguous word *some* varied with the speaker's manner of speech. They set up a social context where eating more cookies is associated with face loss and observed listeners' eye and mouse movements towards plate images with different numbers of cookies left when hearing utterances with ("I ate, uh, some oreos") or without fillers ("I ate some oreos"). The results showed that when hearing disfluencies in this specific context, listeners were fast in deciding and choosing a corresponding inference of ambiguous *some* ("I ate some but not all oreos"). This fast-reasoning process indicated that listeners are able to use disfluencies as a social cue facilitating understanding, bearing the context of speaker's goal of face-saving in mind.

Ambiguity problem is another pervasive phenomenon in daily speech. Ambiguous utterances and how disfluency is connected to disambiguating utterances are discussed further in Chapter 3.

2.4.4 The time-course of comprehension

Besides the possible accounts of how listeners comprehend disfluencies, the time-course of the comprehension process is also a focus when exploring the effects of disfluencies in language processing.

Disfluencies are found to influence the offline *outcome* of how people understand an utterance (Bonnefon, Dahl, & Holtgraves, 2015). Bonnefon et al. (2015) used an offline rating task, showing that disfluencies influence the eventual interpretation of *some* in a given social context. By manipulating the presence of silent pauses in different social scenarios and asking participants to read sentences, such as "some people hated your idea" or "some people loved your idea", Bonnefon et al. found that disfluencies were related to negative meaning of *some* in both face-threatening situations and in face-boosting situations. However, in face-threatening situations, the ultimate interpretation of *some*[5] is the literal interpretation of *some*, "Possibly all people hated your idea", while in face-boosting situations disfluencies were related to pragmatic interpretation of *some*, "Not all people loved your idea".

However, even when the outcome of understanding is similar, the *process* in online comprehension can be different. Many studies in ASD and non-ASD individuals provided evidence for this point. Both groups showed similar results in tasks exploring disfluency comprehension, such as providing context-appropriate explanations for disfluencies in stories (Happé's Strange Stories Test: Happé, 1994), despite the expectation that ASD groups would exhibit less developed abilities in pragmatic reasoning (Hala, Pexman, & Glenwright, 2007; Happé, 1995; Jolliffe & Baron-Cohen, 2000). Visual world experiment in Bailey and Ferreira (2007) also indicated the importance of observing the process of comprehending, as by observing the outcome of participants' actions alone, they would notice that participants were able to correctly follow the directions in either early or late disfluency condition, but only by observing the online process of how participants' fixation towards objects change could they conclude an effect of disfluency in disambiguating syntactic structure.

With evidence from these studies, I emphasise the necessity of looking at the real-time process of listeners' comprehension of disfluencies beyond the outcomes of interpretation. Studies from Loy et al. (2017, 2018, 2019) and King et al. (2017) have also drawn attention to the online effect of disfluencies in real-time language

[5] For detailed discussion of the interpretations of the scalar quantifier *some*, please check Chapter 3.

processing. Eye-tracking and mouse-tracking results from these studies showed that listeners integrate the pragmatic cues, such as disfluencies, quite quickly from early stages of comprehension before the speaker unfolding utterances. Similar results that listeners make pragmatic inferences rapidly in comprehension have also been found in Grodner and Sedivy (2011), Hagoort and colleagues (Hagoort, Hald, Bastiaansen, & Petersson, 2004), Kurumada and colleagues (Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014) and etc.

2.4.5 Summary: the comprehension of disfluencies

Despite being defined as simple interruptions of speech flow, disfluencies are shown to facilitate listeners' language comprehension to reach an ultimate interpretation. There has been evidence that disfluencies can act as pragmatic cues of deception and to heighten attention of the following utterances, or cues for disambiguating speech. Previous studies have shown that the pragmatic cues involved in disfluencies affect listeners' comprehension via a perspective-taking process or via a heuristic. However, with the findings that the effects of disfluencies in listeners' comprehension were context-specific, it remained unclear which account can best explain the question of how disfluencies affect listeners' real-time processing. This thesis aims to distinguish the perspective-taking and heuristic account. Importantly, this thesis will propose a social reasoning account that disfluencies cue a social reasoning process where listeners take the speaker's communication goal in mind and consider the pragmatic functions of disfluencies in the given contexts. With many studies providing evidence for the incremental effects of disfluencies in online language processing, I also note in this thesis the importance of exploring the *process* of how disfluencies exert influence on listeners' comprehension, rather than only focusing on the outcome of comprehension.

2.5 Conclusion: Disfluency in spontaneous speech

Disfluency is a pervasive phenomenon in daily utterances and therefore an important part to understand the production and comprehension of spontaneous speech. The production of disfluency is not arbitrary, as disfluencies are found signalling production difficulties, and following discourse-new, unfamiliar, or low-frequency words. Under communicative circumstances, disfluencies are produced containing communicative information, such as showing intention to keep talking to hold the floor of conversations. When given certain social contexts as hiding deceptional intention to gain rewards, speakers were also found to manipulate the use of disfluencies to make themselves sound more or less trustworthy.

With disfluency defined as interruptions in speech flow, it may be thought to add ambiguity in the process of comprehending, but contextual information actually allows disfluencies to benefit rather than hinder listeners' language comprehension. Disfluencies' effects on helping comprehension process manifest as acting as pragmatic cues which lead listeners to a social reasoning process and disambiguate utterances in a specific context. This thesis, on the basis of previous studies on exploring the pragmatic role of disfluencies, focuses on investigating the specific social reasoning process of disambiguating utterances and its time-course behind listeners' uses of disfluencies in social contexts. To further discuss how disfluencies may provide resolution for ambiguous utterances in language comprehension, I describe what ambiguous utterances are and elaborate the relationship between disfluencies and ambiguity in spontaneous speech in Chapter 3.

Chapter 3
Pragmatic Ambiguity in Speech

3.1 Introduction

Following the discussion of disfluency's role in language comprehension, this chapter reviews the ambiguity literature that is specifically related to the aforementioned research question. In this thesis, I specifically focus on pragmatic ambiguity in spontaneous speech. To review prior work on ambiguity resolution in language comprehension, I first introduce two major language processing models. With a special focus on pragmatic ambiguity, I define ambiguous utterances in the background of this thesis. I draw special attention to a specific case of ambiguity, the scalar quantifier *some*, discussing how *some* can be ambiguous in utterances. The focus of this chapter is how listeners process the scalar quantifier *some* in speech with disfluencies, in which the time-course of the processing and the influence contexts draw on listeners' processing are elaborated. Combining the evidence from previous studies, I highlight disfluency's pragmatic role in the facilitation of understanding *some* in real-time comprehension and pinpoint the unanswered question on how disfluencies exert influence on listeners' processing of *some*, specifically how listeners integrate the pragmatic cues from disfluencies with a given social context in which the utterance is produced.

It is important to note that the studies and papers discussed in this chapter are selected to illustrate key concepts and mechanisms relevant to the ambiguity resolution or disfluency processing. These examples serve as representative cases rather than exhaustive or definitive statements on these topics.

3.2 Language processing models

A successful understanding of utterances requires listeners to resolve the ambiguous parts in speech. Previously, different models of language comprehension have been proposed to explain how listeners resolve temporary ambiguities encountered in processing. One type of comprehension model assumes a serial computation, such as garden-path model (Frazier, 1987; Frazier & Rayner, 1982). These models are serial, meaning that they produce a single interpretation at a time. Specifically, the garden-path model is modular in nature, prioritising the role of syntactic category information and indicating that the initial processing is isolated from other types of knowledge. The garden-path model only uses all available information in the second stage to check whether the first-stage analysis is plausible and initiate a re-analysis of the structure if necessary.

This model has not gone unchallenged by observations that factors beyond sentence structure alone affect sentence processing, leading to the development of the constraint-based model (McClelland, St. John, & Taraban, 1989; Spivey-Knowlton & Sedivy, 1995; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993). Different from serial models like the garden-path model, constraint-based models assume a parallel and non-modular computation. In constraint-based models, the current interpretation reflects the integration of multiple competing constraints, meaning that various sources of linguistic and non-linguistic information are gradually integrated into the comprehension system to compute the most plausible interpretation. Importantly, constraint-based models suggest that different types of information continuously interact until the listener settles on a final interpretation. Under the constraint-based model, the comprehension of utterances is context-driven and this model suggests that listeners integrate multiple sources of information as soon as the utterance unfolds.

Bearing in mind this brief introduction of two major language comprehension models proposed by previous studies, I review ambiguity literature in the following sections, with a special focus on pragmatic ambiguity.

3.3 What is pragmatic ambiguity?

Ambiguity is an inherent feature of human language – far from being an obstacle to communication, it reflects a dynamic interplay between a speaker's choices of words and the complexity of listeners' experience and perspectives. In the simplest sense, one reason that ambiguity arises in language is because reality is much more complex than language (Pehar, 2001). Ambiguity resolution is therefore a core task in language comprehension.

While ambiguity is a pervasive phenomenon in both written and spoken language, I focus in this thesis specifically on the latter type – ambiguous utterances. By focusing on utterances, I emphasise two features. The first feature is that utterances involve both sides of interlocutors (i.e., speakers and listeners). The second feature, demonstrating the additional difficulty of speech compared to written language, is that ambiguities in utterances must be resolved in real-time to achieve a communicative goal.

One useful definition of an ambiguous utterance is that it involves "interpretive uncertainty" (Kennedy, 2009), meaning that multiple meanings can be mapped to one utterance. To detail this definition in terms of this thesis, "interpretive" represents a listeners' perspective, and "uncertainty" specifically refers to the uncertainty of assigning a distinct meaning to an expression which can be understood in different ways, perhaps according to context. Despite this definition being general, it provides a good basis for the exploration of the phenomena in human spoken language outlined in this thesis.

Defined in such a way, daily speech can be ambiguous at all levels of linguistic analysis (i.e., lexical, syntactic or pragmatic). For the purposes of introducing the concepts that are of the most interest for this thesis, I give special focus to pragmatic ambiguity. In the next section, I define pragmatic ambiguity and explain why this specific ambiguity is of focus in this thesis.

3.3.1 Define pragmatic ambiguity

Pragmatic ambiguity often arises in a communicative environment, when an utterance can be interpreted in multiple ways based on contexts and implicatures drawn from the conversations. It is not the words or sentence structures that are ambiguous, but rather the intention, belief and attitudes of a speaker that inform meaning in a given context.

(1) Can you open the window?

For example, sentence (1) does not contain any ambiguity in the sentence itself, but it could be interpreted in various ways if considering this sentence in different contexts. This sentence could be a request, asking the listener to open the window, or it could also mean that the speaker would like to offer someone help by asking if that person is capable of opening the window.

Among all types of ambiguity, pragmatic ambiguity is unique in three ways. First, pragmatic ambiguity arises not only from the properties of words, but also from the broader context of utterance exchanges. Its resolution involves the listener's consideration of the speaker's intention and beliefs, which adds complexity beyond what is typically seen with lexical and syntactic ambiguities. Second, unlike other ambiguities that often require a precise resolution for comprehension (e.g., determining the meaning of the word "bank" to resolve lexical ambiguity), pragmatic ambiguity does not always necessitate such precision. For example, when a speaker says, "It's getting late", it could be interpreted as a suggestion to leave rather than merely an observation about the time. This kind of ambiguity often remains unresolved to a unique and agreed-upon meaning, as speakers can be deliberately vague and listeners are sensitive to this vagueness. Third, the words involved in pragmatic ambiguity do not always have explicitly distinct meanings, making this type of ambiguity more susceptible to social contexts. In these fronts, pragmatic ambiguity is dynamic and context-sensitive, which makes it a critical area of focus for understanding language comprehension.

Conversational implicature is an important cause of pragmatic ambiguity in utterances – whenever there arises conversational implicature, the utterances are said to be pragmatically ambiguous (Francesch & Payrató, 2024). In the rest of this section, I explain in detail the concept of conversational implicature and the causes of pragmatic ambiguity due to the occurrences of conversational implicatures, as well as a specific case of this kind of pragmatic ambiguity.

3.3.2 Conversational implicature and its causes

Conversational implicatures (CIs) are part of the content implied by the speaker in making an utterance, additional to the literal meaning of a given utterance (Blome-Tillmann, 2013; Francesch & Payrató, 2024). Under this definition, CIs are regarded as independent of the conventional meaning of the utterances but determined largely by the context of the uttered speech (Blome-Tillmann, 2013).

According to Grice (1975) and Levinson (2000), CIs are related to the Maxims of Conversation, which form a prerequisite for communication assuming that speakers follow a cooperation principle in conversing. If following the maxims of conversation, speakers are expected to adjust the amount of information provided according to conversational needs (Maxim of Quantity), to provide accurate information and truths in conversations (Maxim of Quality) and to utter information that is relevant and related to the conversation topic (Maxim of Relevance).

In turn-taking, speakers following or flouting these maxims of conversation cause listeners to draw various kinds of conversational implicatures, leading to eventual pragmatic ambiguity. For example, if someone asks, “Do you know where the library is?” and the response is “It’s somewhere on campus”, the speaker may be flouting the Maxim of Quantity by providing insufficient information or the Maxim of Quality by being vague. This response can lead the listener to draw a conversational implicature that the speaker either does not know the exact location or is being deliberately unhelpful, thus creating pragmatic ambiguity.

A special case of this kind of pragmatic ambiguity arises from the interpretation of Scalar Quantifiers, a term referring to words such as *some*, *a few*, *many*, etc. For the topic this thesis concerns, I draw special attention on the scalar quantifier, *some*, and briefly introduce in the following section how *some* can be ambiguous in utterances. I discuss listeners’ processing of *some* in section 3.4.

3.3.3 A specific case: Scalar quantifier, *some*

The scalar quantifier *some* can be ambiguous between two elicited meanings, a literal interpretation (2b) and a pragmatically strengthened meaning (2c). See the sentences in (2) for an example:

(2) a. Some students got ‘A’s.
b. Some, and possibly all students, got ‘A’s.
c. Some, but not all students, got ‘A’s.

Sentence (2)a can yield a literal interpretation of (2)b due to the lower-bound meaning of *some* being “at least one”. The pragmatic interpretation (2)c of *some* to be “not all” is called Scalar Implicature, and it is regarded to be a pragmatic reasoning of listeners considering Gricean maxim of quantity (Grice, 1975): If the speaker is cooperative and knows that all students got ‘A’s, then they are expected to utter *all* instead. But the fact that the speaker uses *some* indicates that they are not in the position to utter *all*, a stronger scalar quantifier. This reasoning process results in the scalar implicature “some but not all”.

Ambiguity between these two interpretations of *some* has led to exploration of extensive research topics, such as how listeners resolve this ambiguity and reach a final interpretation of this ambiguous word, as well as the time-course of this reasoning process of listeners’ comprehension. With the concepts introduced in the preceding sections, I focus on discussing how listeners resolve pragmatic ambiguity in language processing in the following section.

3.3.4 Summary: causes of ambiguity in utterances

Spontaneous speech is ambiguous in nature and an utterance can contain ambiguity in all linguistic levels, specifically lexical, syntactic and pragmatic levels. With a focus in pragmatic level of ambiguity in this thesis, evidence from previous studies has shown that one important cause of pragmatic ambiguity in daily speech is conversational implicatures. These implicatures’ occurrences are due to listeners’ pragmatic reasoning about the speaker’s statements or states considering Gricean maxims of conversations (such as the Maxims of Quantity, Quality, Relevance etc.), and their eventual interpretations are largely dependent upon contexts. I introduce the Scalar quantifier *some* as a specific case of this kind of pragmatic ambiguity. The meaning of *some* can be literal or pragmatically strengthened depending on given contextual information. In the next section, I discuss how listeners resolve ambiguity in comprehension with a special focus on the processing of the scalar quantifier *some*.

3.4 How do listeners resolve pragmatic ambiguity in speech?

Pragmatic ambiguity, compared to all the other types of ambiguities, is special as it is predicated on knowledge more than linguistic factors (i.e., lexical frequency or syntactic structure[6]). Contexts are therefore an important factor for listeners to disambiguate utterances. For example, Swaab and colleagues (Swaab, Brown, & Hagoort, 2003) used the N400, an ERP component linked to language processing, to measure how contexts affect listeners' selection of the appropriate meaning of ambiguous words (e.g., bank) in Dutch sentences. In a concordant context, the context biases the meaning of the ambiguous word to be related to the target word. For instance, the sentence "My friend meets with her lawyer"[7] which biases the legal interpretation of the ambiguous target word "judge", serves as a concordant context. In contrast, a sentence like "My mother eats her eggnog" where the contexts does not relate to any plausible meaning of "judge", provides a discordant context. Their results showed that the N400 amplitude was reduced in concordant contexts, indicating facilitated processing. Conversely, the N400 increased in discordant contexts, reflecting increased processing difficulty. The N400 amplitude was largest when the context ended with an unambiguous noun unrelated to the target word, indicating the highest processing difficulty. Their findings suggest that, upon hearing an ambiguous utterance, despite initial activation of both dominant and subordinate meanings, the context gradually guides listeners to select the appropriate meaning of an ambiguous word.

Visual context was also found to facilitate listeners' comprehension of ambiguous utterances (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). When hearing instruction like "Put the apple on the towel in the box", listeners were given visual contexts either supporting the normally preferred interpretation "the apple should be put on the towel" or the less-preferred interpretation of "the apple that is already on the towel should be put in the box". Using eye-tracking equipment, Spivey et al. (2002)

---

[6] Factors such as lexical and syntactic frequencies were found to be related to resolution of lexical and syntactic ambiguities (such as MacDonald, 1993; MacDonald, Pearlmutter, & Seidenberg, 1994; Novick, Kim, & Trueswell, 2003; Spivey & Tanenhaus, 1998; Trueswell, 1996). These factors are not related to the main research question of this thesis, and therefore are not explained in detail.
[7] The example sentences provided here are literal English translations of the original Dutch materials used in Swaab et al. (2003).

observed specific eye-movement patterns which suggest a tendency to initiate an interpretation consistent with the context.

Contextual information affects listeners' processing in real time to facilitate their disambiguation process of comprehension, but so far, few studies have explored the role of social contexts on influencing listeners' comprehension of ambiguous utterances and the time-course of listeners' integration of this contextual information. In the following sections, I review the evidence of previous studies on social contexts' effects on listeners' comprehension, taking utterances including the scalar quantifier *some* as the specific example of ambiguous utterances. I discuss how listeners comprehend the scalar quantifier *some*. I then discuss the role of contexts, especially social contexts in listeners' processing of *some*. Finally, the connection between contextual cues triggered by disfluency (Chapter 2) and how they are utilised by listeners to interpret *some* in spontaneous utterances are discussed.

3.4.1 Interpreting scalar quantifier *some*

As shown in sentence (2), the ambiguity caused by the scalar quantifier *some* involves an interaction process between its semantic and pragmatic meanings. Specifically, the pragmatically strengthened interpretation "some but not all" is called scalar implicature. Scalar implicature occurs when a speaker uses a weaker term (e.g., uses the term [some]) from a scale of elements ordered by their semantic strength (e.g., from the scale [all, most, many, some, few]; see Horn, 1972), listeners make inferences that the stronger term does not apply (e.g., "not all"). Listeners' comprehension of the scalar quantifier *some* is therefore dependent largely on their choices between these two interpretations, resulting in a literal interpretation or deriving a scalar interpretation.

Some accounts assume that listeners' processing of these two meanings of *some* follows a serial computation model, such as literal-first account (Bott & Noveck, 2004), pragmatic-default account (Chierchia, 2004, 2006; Levinson, 2000), and two-step account (Tomlinson Jr., Bailey, & Bott, 2013). These accounts predict either (i) that the literal meaning of *some* is derived first and listeners enrich this meaning to infer the implicature when assuming the speaker is informative and cooperative, or (ii) that the pragmatic implicature of *some* is stored in the lexicon and derived automatically.

Other accounts propose that listeners' comprehension of *some* aligns with the constraint-based model, where listeners consider multiple cues to determine whether a scalar implicature is likely (Breheny, Katsos, & Williams, 2006; Degen & Tanenhaus, 2014). These accounts suggest that listeners' processing is influenced by cues from various available contexts, including linguistic, discourse and global contexts relevant to the comprehension goal.

These models fundamentally differ in their predictions about the time course of processing *some* (see Chemla and Singh, 2014a, b for reviews), as well as how and when context plays a role in processing. The following subsections discuss these two perspectives in detail, focusing on the time course of interpreting *some* and the influence of contexts. In this thesis, the focus is on the ways in which various types of contexts play a role in listeners' real-time processing of *some*, highlighting the implications for the models mentioned above.

3.4.1.1 Time-course of interpreting some

Previous studies observed differences in the time course of listeners' deriving the pragmatic implicature interpretation of *some* which led to various accounts of the processing of *some*. Individuals' intuitive ease of interpreting *some* to imply "not all" has prompted accounts proposing that the pragmatic implicature interpretation is linked to lexicons and is therefore computed automatically via a specific grammatical system (Chierchia, 2004, 2006; Levinson, 2000).

However, listeners' access to the scalar implicature interpretation of *some* was found in some studies to be slow and cognitively demanding. Specifically, a processing difficulty for the pragmatic meaning of *some* has been observed in self-paced reading and sentence verification tasks: participants were found to take longer to respond (Bott & Noveck, 2004; Noveck & Posada, 2003) or read sentences (Bergen & Grodner, 2012; Breheny et al., 2006) with the scalar interpreted as "some but not all" compared to its literal counterpart, and their accuracy decreased in sentence verification tasks requiring quick responses (Bott, Bailey, & Grodner, 2012). In the sentence verification tasks of Bott and Noveck (2004) and Noveck and Posada (2003), researchers found that when given ambiguous sentences like "Some elephants are mammals", participants took longer to evaluate the sentence when they were instructed to use the pragmatic interpretation ("Some but not all elephants are mammals") compared to when instructed to use the literal interpretation ("Some and possibly all elephants are mammals"). Moreover, in the unrestricted condition when participants were not given any instructions of interpreting *some*, those who intuitively chose the pragmatic interpretation took longer than those choosing the literal interpretation. These results argue against pragmatic-default accounts, and in favour of accounts which suggest that deriving the literal meaning is faster and deriving the scalar implicature interpretation is effortful with deeper processing.

Consistent with the findings from these tasks, in an online eye-tracking study, Huang and Snedeker (2009) observed a delayed fixation towards the referent when participants need to compute a scalar implicature. In their visual world paradigm, participants were asked to follow audio instructions like "Point to the girl that has *some/two/all/three* of the ice cream sandwiches". Initially, equal fixations were found on both the literal referent (the girl who has all of the ice cream sandwiches) and the pragmatic referent (the girl who has a subset of the ice cream sandwiches). Fixations towards the pragmatic referent (the girl with a subset) only exceeded the chance level from 1100 milliseconds after the onset of some.

Results from these studies correspond to a delay in processing time for the pragmatic meaning of *some*, indicating an effortful pragmatic strengthen on the basis of the initial literal interpretation of *some*. Critically, the observed delay in accessing the scalar implicature supports an account that pragmatic enrichment is a secondary process rather than an immediate, automatic response. These findings therefore suggest an account aligning with modular, sequential models of processing, such as literal-first or two-step account.

However, different results from similar experimental paradigms have challenged this position – Grodner and colleagues (Grodner, Klein, Carbary, & Tanenhaus, 2010) used a similar visual world paradigm to Huang and Snedeker (2009), but they failed to find delayed fixations towards the pragmatic implicature meaning. Instead, the pragmatic interpretation seemed to emerge from the earliest moment of comprehension.

One reason they proposed for this different result was regarding the appropriateness of the given contexts. In Huang and Snedeker's experimental setting, the existence of filler sentence with exact number, such as "Point to the girl with two of …", could have made the instruction sentence with the critical word *some* sound less natural when the target image also includes two objects. The delayed fixation time observed was possibly not due to processing scalar implicature itself, but rather because of having difficulty in integrating the available contextual information into processing. In this case where the target image included two objects, the difficulty might arise when listeners consider the alternatives the instruction could have used, and they would take this information with context to infer why *some* rather than the exact number *two* was used (discussed also in Degen & Tanenhaus, 2014). When Grodner et al. removed trials with exact numbers, they found that listeners were equally fast in interpreting *some* in its pragmatic meaning and literal meaning, both focusing the target image 200-300ms after the onset of *some*, at an early stage of comprehension.

Grodner et al.'s findings highlight the critical role of context in processing the scalar quantifier *some*. Their results demonstrate that listeners are able to quickly infer the pragmatic meaning of *some* when provided with a clear context without competing alternatives. These findings align with a constraint-based model, suggesting that pragmatic constraints can influence comprehension from the earliest moments. This conclusion is also supported by other studies indicating that listeners can quickly draw pragmatic inferences either about the speaker or their utterances in the early stages of comprehension (e.g., Grodner & Sedivy, 2011; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; Van Berkum, 2009; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008).

In the section below, I further review the evidence from studies discussing how contexts exert influence on listeners' processing of the scalar quantifier *some*.

3.4.1.2 Contexts and the comprehension of some

As shown in the results of Grodner et al. (2010), the contextual sensitivity in the processing of scalar quantifier *some* has also been forefronted by many other researchers. One type of the contexts relevant to this comprehension process is the discourse context (e.g., Politzer-Ahles & Fiorentino, 2013; Politzer-Ahles & Husband, 2018; Breheny et al., 2006). Consider the following two dialogues:

(3) a. – How many students have passed the exam?
– *Some* of the students passed.
And *the rest* are going to retake the course.

b. – Have *all* the students passed the exam?

– *Some* of the students passed.

And *the rest* are going to retake the course.

In (3)a, the word *some* is ambiguous as the person answering is assumed to either know the fact and uses *some* to express the meaning that "some but not all students passed the exam", or be unsure about the fact and uses the word to presume "some and possibly all students passed the exam". Compared to (3)a, question in (3)b evidently focuses on whether *all* is true, which provides a more convenient discourse context for comprehenders to interpret the use of *some* pragmatically, as a negation of *all*. Subsequently, with the "not all" interpretation of *some*, comprehenders are expected to be more aware of the remaining subset of the students and process "the rest" part more quickly in (3)b than (3)a condition. Using self-paced reading, several studies have verified this assumption, with the result of a faster reading times over the phrase "the rest" in (3)b (Breheny et al., 2006; Katsos, 2007; Panizza, Chierchia, & Clifton, 2009). Moreover, on the basis of manipulating the discourse context of whether *all* was made salient in the preceding turn, Politzer-Ahles and Husband (2018) added an *only some* control condition (i.e., *only some* of the students passed the exams) in which the interpretation of *some* is not dependent on context. They replicated the faster reading time for *the rest* in a context supportive of a pragmatic implicature interpretation (as (5)b), but also observed that this pattern attenuated significantly in *only some* condition which is a semantically fixed condition. These findings suggest that the context influences the comprehension of *some* depending on whether the discourse context supports a literal or a pragmatic interpretation.

Using a similar paradigm, previous studies have also indicated that how quickly *some* is processed depends on the comprehenders' knowledge of the broader context of the discourse or the speaker's states (Bergen & Grodner, 2012; Breheny, Ferguson, & Katsos, 2013). By comparing participants' reading times at each part of the given sentences (e.g., Some of/the real estate/investments lost; The rest/were successful/despite the/recent economic downturn), Bergen and Grodner found that in the cases where the speaker was thought to know whether the stronger *all* statement is true, participants read the sentences with *some* slower, but processing *the rest* faster, compared to in the cases where they thought the speaker merely might know the stronger statement is true. This result replicated the processing difficulty of the pragmatic meaning of *some* and the point that the realisation of implicature interpretation facilitates the processing of *the rest*. Critically, this study highlighted the role of context influencing the processing speed of *some*, mediated by incrementally taking the discourse or the speaker's state knowledge into consideration.

Beyond linguistic contexts, many literatures have also indicated the interpersonal properties of *some*, such as serving a politeness function (Bonnefon, Feeney, & Villejoubert, 2009; Bonnefon, Neys, & Feeney, 2011; Holtgraves & Kraus, 2018). In an off-line task, Bonnefon et al. (2009) asked participants to read an utterance with a given scenario and provide a judgement of how possible the *some* in that utterance guarantees a literal interpretation ("some and possibly all…"). They found that when the given scenario is face-threatening (e.g., *Some people hated your poem*), participants showed more tendency to endorse the possibility of a literal interpretation ("Some and possibly all people hated your poem"), while in a face-boosting scenario (e.g., *Some people loved your poem*), participants favoured more of a pragmatic interpretation ("Not all people loved your poem"). When analysing participants' response times for interpreting *some* in various scenarios, Bonnefon et al. (2011) found that in the face-threat scenarios, those who endorsed a literal interpretation took significantly longer to respond than those who did not. This result suggests that face-threat contexts promote the literal interpretation of *some*, but make it effortful and time-consuming to achieve. Previous studies (e.g., Breheny et al., 2006) suggested that contexts frequently leading to a literal interpretation of *some* also made this interpretation easier for people to process. However, Bonnefon et al. (2011) demonstrated an opposite effect in contexts with a politeness goal involved, indicating that comprehenders may face extra processing demands beyond the usual process of interpreting *some* in this given context.

Holtgraves and Kraus (2018) further tested the effect of face-threat using a self-paced reading task with a measure of participants' neural responses (ERP). Participants were provided with natural conversational scenarios with different levels of face-threat, followed by a target utterance with the scalar quantifier *some* and a continuation utterance aligning either with a literal interpretation or a pragmatic interpretation. A higher P300[8] was observed when the continuation utterance used the

[8] P300 is an ERP component that is often associated with attention and the processing of unexpected or particularly significant stimuli.

literal meaning, and this difference of P300 between a literal or pragmatic interpretation continuation utterance was more pronounced in scenarios with higher levels of face-threat. This result suggested that in such conversational contexts concerning maintaining face, participants initially anticipated a pragmatic interpretation and therefore were surprised when seeing a continuation sentence with the literal meaning. These studies altogether suggested that *some* can be taken by listeners as a politeness device to mitigate a potential face-threatening scenario, emphasising that the processing of *some* is influenced by the social contexts of conversations.

Besides the global contexts in which the scalar quantifier *some* occurs, the speaker's manner of speech has shown to be a local factor influencing listeners' processing of *some*. I discuss this claim in the following section to pinpoint the role of disfluencies (discussed in Chapter 2) in affecting the interpretation of scalar quantifier *some* in utterances.

3.4.2 Disfluency and *some*

As discussed in Chapter 2, the speaker's manner of speech (in this thesis specifically, whether the speaker's utterances contain a disfluent filler *uh* or not) is shown to play a communicative role, in that listeners are able to take it the speaker's disfluency as a pragmatic cue of deception or a cue to heighten attention. Instead of hindering the comprehension and adding ambiguity to the utterances, the pragmatic function of disfluency has been suggested to facilitate comprehending ambiguous utterances via a social reasoning process (as mentioned in Chapter 2). Moreover, disfluency has been found to influence listeners' process of making pragmatic inferences and listeners integrate the paralinguistic cues from disfluency from early stages of comprehension which shape their ultimate utterance interpretation (as discussed in Chapter 2).

With these pragmatic features of disfluency and the fact that interpretations of *some* depend on the social contexts (detailed in previous sections), there are reasons to believe that the speaker's manner of speech can influence listeners' interpretation of the specific case of ambiguous scalar quantifier *some*. Few empirical studies have tested this effect of disfluency in listeners' processing of some (Bonnefon, Dahl, & Holtgraves, 2015; Loy et al., 2019). On the basis of previous studies suggesting that the comprehension of *some* is dependent on the speaker's politeness goals involved in maintaining-face contexts, Bonnefon et al. (2015) further tested the effect of disfluency (specifically silent pauses) in such contexts. They provided written scenarios in which the description of whether the speaker had a silent pause before uttering a face-threatening statement was manipulated (such as sentence (4)).

(4) You asked what Bob thought about your idea.
Bob <stays silent for a few seconds. Then he> replies:
"Some people hated your idea." *(face-threatening statement)*
Or "Some people loved your idea" *(face-boosting statement)*

Participants were asked to rate how possible the statement reaches a literal interpretation of *some*. Bonnefon et al. found that regardless of whether it was in a face-threatening or a face-boosting condition, when the speaker was described as remaining silent before replying, participants favoured the unpleasant interpretation specific to that context – a literal interpretation in face-threatening situation ("Possibly all people hated your idea") and a pragmatic interpretation in face-boosting situation ("Not all people loved your idea"). While Bonnefon et al. (2011) showed that interpreting *some* is effortful as people took long time to decide whether to interpret *some* literally or pragmatically in face-threatening contexts, results from Bonnefon et al. (2015) suggested that disfluency facilitated the comprehension of *some* by functioning as a social cue. Bonnefon et al. (2015) interpreted this result that disfluency signalled to comprehenders that the upcoming statement was unpleasant, thereby prompting an unpleasant interpretation of the statement. Another way to interpret this result is that comprehenders might reason about the speaker's disfluency, considering whether the speaker is taking longer (i.e., staying silent) to formulate a polite yet truthful response. However, the use of written utterances in the study by Bonnefon et al. (2015) limits the exploration of this potential reasoning process cued by disfluency, as reasoning about written scenarios might not be equivalent reasoning about spoken utterances.

While Bonnefon et al. (2015) found an off-line effect of disfluency on the eventual interpretation of the scalar quantifier *some*, Loy et al. (2019), using eye- and mouse-tracking methods, tested how disfluency affects the online processing of utterances with *some*. Loy et al. created a social context similar to that of Bonnefon et al. (2015) which is related to face-loss concerns, in which the literal meaning of some

("some and possibly all…") is the interpretation threatening the good image of the speaker. They provided a cover story to achieve this, describing a set-up experiment exploring the relationship between individuals' snack habits and greediness. The cover story described a scenario where a group of people were allowed to eat various snacks while watching a documentary and they could take any wanted number of snacks. However, the real purpose of investigating greed was revealed only after the documentary finished and then each individual was asked to report how many snacks they had eaten.

Participants in Loy et al's studies were told that they would hear recordings of these individuals' answers. A social context related to face-maintaining was set up in this way, as speakers who had taken *all* snacks could purposely use the ambiguity of *some* to avoid being judged to be greedy and disfluency could be a by-product of this effortful calculation process. Simply put, Loy et al. expected a disfluent answer, such as "I ate, *uh, some* oreos" to suggest to the listener that this speaker ate *all* of oreos but was unwilling to admit it. In the experiment, recorded sentences with manipulation of speakers' manner of speech (disfluent/fluent) were played while participants were presented a visual display of two plates containing the remaining quantity of snack items – one was an empty plate, corresponding to a literal interpretation ("I ate all of the [snack]"); while the other plate with a number of remaining snacks corresponded to a pragmatic meaning ("I ate some but not all of the [snack]"). In each trial, participants' eye and mouse movements were recorded while choosing to click one of the plates.

Loy et al. (2019) found that participants favoured a pragmatic interpretation of *some* ("I ate some but not all oreos") hearing fluent utterances, but disfluent utterances led participants to favour a literal interpretation instead ("I ate some and possibly all oreos"). Moreover, this difference between conditions emerged in the eye- and mouse-tracking data as soon as the offset of *some*, once again showing listeners' early integration of pragmatic inferences in comprehension (as in King et al., 2017; Loy et al., 2017). Results from this study are indicative on two fronts: listeners' processing of *some* is context-dependent, in that disfluency led listeners to favour an unpleasant interpretation in this potential face-loss context; and listeners make rapid use of disfluencies to make inferences of *some* in real-time comprehension.

With the use of spoken utterances, the findings of Loy et al. (2019) build upon those of Bonnefon et al. (2015), providing further evidence of disfluency's role as a social cue and its rapid effect on the real-time comprehension of ambiguous *some*. However, by focusing solely on face-threatening scenarios, Loy et al. leave several questions unanswered. For example, considering the accounts of listeners' processing of disfluency discussed in Chapter 2, is disfluency merely a heuristic cue indicating unpleasant messages, or does it serve as a social cue prompting listeners to reason about the speaker's communication goals when encountering disfluent utterances with *some*? If disfluency is related to a reasoning process, how is it manifested in listeners' interpretation of *some*? Are listeners processing disfluency as a cue that heightens their attention to the basic meaning of *some*, or as a cue that leads them to question the speaker's truthfulness?

The empirical work in this thesis detailed in Chapters 5-6, focuses on exploring further on these questions.

3.5 Conclusion: Disfluency in influencing the interpretation of *some*

The scalar quantifier *some* is ambiguous between a literal interpretation ("some and possibly all") and a pragmatic strengthened interpretation ("some but not all"). Previous studies exploring the processing of *some* indicated that the disambiguation of this word depends on context. Results from both off-line and real-time tasks have shown that discourse contexts (i.e., whether *some* occurs in a context supporting pragmatically strengthened meaning), as well as contexts of the speaker's knowledge states influence people's speed and outcome of interpretation. Moreover, social contexts, such as a context involving the speaker's politeness goal, have also shown influence on listeners' real-time interpretations of *some*. Listeners have been found to make rapid use of pragmatic inferences drawn from the social context and reach a final interpretation via reasoning processes. One social cue listeners use is the speaker's manner of speech, specifically, whether the speaker utters a filler with *some* (*uh, some…*). Similar to discussion of disfluency's role in language comprehension in Chapter 2, evidence has shown that disfluencies function as pragmatic cues influencing listeners' real-time processing of *some*. However, previous studies leave open questions regarding specifically how listeners make use of disfluencies in real-time processing, such as what reasoning process is cued by disfluencies. For example, whether listeners regard disfluencies as a cue to heighten listeners' attention to the following words or a cue of potential deception, and how these reasoning processes are manifested in their real-time comprehension.

This thesis focuses on exploring the pragmatic role disfluency plays in real-time processing, by combining an investigation of how disfluency interacts with social contexts to resolve ambiguous utterances, with a further exploration on the individual differences in pragmatic ability manifested in comprehending disfluent utterances. To address the limitations in contexts provided in previous studies, an attempt was made by varying social contexts in tasks. To keep the track of exploring the role of disfluency in real-time processing, methods of eye-tracking and mouse-tracking are applied. Both techniques are informative in terms of the behavioural process during language processing, which make them suitable to observe any process difference caused by disfluency and contexts. Before presenting the experimental work (Chapter 5-7), I describe the methodologies of eye-tracking and mouse-tracking, as well as their application in exploring questions about language comprehension.

Chapter 4
Mouse-tracking and Language Comprehension

4.1 Introduction

Mouse-tracking and eye-tracking are two methodologies applied in this thesis as they provide a window into real-time language processing. Eye-tracking is a commonly used technique for comprehension tasks in psycholinguistic studies, which has been discussed extensively (see Boland, 2004; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Huettig & Altmann, 2005; Tanenhaus & Spivey-Knowlton, 1996; Tanenhaus & Trueswell, 2006 for reviews). Eye-tracking was used in one study only in this thesis and its use is specified in Chapter 7. Mouse-tracking, however, as a method that also has the potential to be efficient in investigating the dynamics of people's cognitive processes, has been less employed in psycholinguistic studies and needs more attention.

Mouse-tracking is valuable for studying language processing as it can track competition between interpretations in real time. Previous studies that have tested for competition between interpretations at different levels, such as the phonological level (Spivey, Grosjean, & Knoblich, 2005) and the semantic/pragmatic level (Tomlinson Jr., Bailey, & Bott, 2013).

In this chapter, I focus on the use of mouse-tracking in psycholinguistic studies. I explain why mouse-tracking is an appropriate method to explore the research questions raised in this thesis. I introduce a typical two-item mouse-tracking paradigm and review the use of this paradigm in the previous literature. I summarise three common ways of analysing mouse-tracking data regarding their advantages and disadvantages. I then introduce a four-item mouse-tracking paradigm specifically designed for the studies in this thesis. I further explain the specific analysis method used for the current mouse-tracking studies and describe a standardised way of reporting mouse-tracking results, which I will employ in Chapters 5-6.

4.2 Mouse movements in revealing real-time language comprehension

In recent years, mouse-tracking has become a popular method in psycholinguistics for providing insights into the dynamic cognitive processes, such as decision making (e.g., McKinstry, Dale, & Spivey, 2008) or language processing (e.g., Farmer, Cargill, Hindy, Dale, & Spivey, 2007; Spivey et al., 2005; Tomlinson Jr. et al., 2013). This method is able to capture the dynamics of mouse movements as participants interact with visual and audio stimuli, revealing the temporal unfolding of mental processes (Freeman & Ambady, 2010; Spivey et al., 2005). Despite their promise, mouse-tracking methods have not been as extensively utilised in language comprehension studies as eye-tracking.

Mouse-tracking was chosen as the main method in this thesis for the following three reasons. The first reason is its effectiveness of unveiling the language processing in real time. The second is the convenience of implementing mouse-tracking paradigms and conducting experiments (either web-based or in-lab). The third lies in the fact that most experiments were designed during the pandemic. Web-based mouse-tracking tasks, therefore, served as an appropriate alternative method to in-person eye-tracking, ensuring the continuation of research during that time.

A critical benefit of the mouse-tracking method in exploring real-time language comprehension is that it records a continuous hand movement response (therefore, smooth mouse trajectories) during the process of comprehension tasks. Eye-tracking has also been regarded as a useful method for studying incremental processing during language comprehension. However, eye movements involve ballistic movements. Saccades are rapid and typically fixed in their trajectory and inalterable once initiated. In contrast to saccades, mouse movements[9] are not ballistic and can be adjusted while they are in progress (Farmer et al., 2007; Goodale, Pelisson, & Prablanc, 1986;

[9] I am aware that there might exist device differences for participants taking part in mouse-tracking studies (mouse or touchpad). However, It is not the focus of this thesis to differentiate these mouse device differences on task performances. Therefore, in this thesis, I generally term participants' motions recorded in mouse-tracking tasks as "mouse movements" for the convenience of description.

Spivey et al., 2005). Therefore, by tracking continuous mouse movements, the path of the mouse is able to reveal the influence of competing stimuli whereby a comprehender's mouse trajectory is pulled or diverted from a competing image stimulus, besides recording the object that is eventually reached.

Nevertheless, one point noted in Spivey et al. (2005) is that it would be incorrect to regard the continuous mouse-tracking data as strictly better than ballistic eye-tracking data. These two methods are complementary in that mouse movements are better at showing **how** different perceptions of speech compete between conditions over time, eye movements are better at showing **when** the perceptions start to activate as saccades are highly sensitive to the timing of early partially formed representations (Spivey et al., 2005). Therefore, the most advantageous way should be using these two methods simultaneously (e.g., Loy, Rohde, & Corley, 2017). A combined use of eye-tracking and mouse-tracking methods had been the original plan for the studies in this thesis, but as explained previously, due to the pandemic, these two methods were used independently in this thesis, with four mouse-tracking experiments and one eye-tracking experiment.

Below I describe what a basic mouse-tracking paradigm is like and summarise several previous studies that have motivated this paradigm for studying listeners' language comprehension. I then briefly outline three commonly used methods for analysing eye-/mouse-tracking data.

4.2.1 Basic mouse-tracking paradigm

A typical mouse-tracking experiment involves a repeated task where participants view a visual scene consisting of two images and make a decision about which image to click on. In these tasks, a participant listens to audio stimuli and moves the mouse cursor from a starting point at the bottom centre of the computer screen, with two visual stimuli as options positioned in the top left and top right of the screen (as shown in Figure 4.1). The experimental software records the participants' final decision of which image to click, as well as the streaming x- and y-coordinates of the mouse cursor. Visual stimuli acting as response options vary depending on tasks and reflect the research questions of a given study.

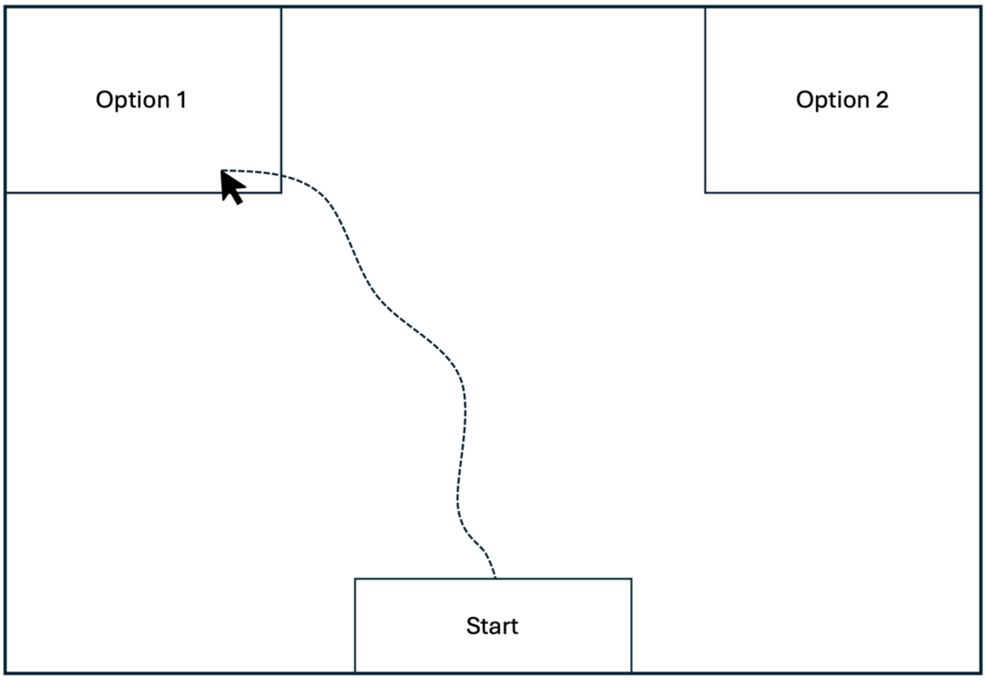


Figure 4.1 An example visual display of a basic two-item mouse-tracking paradigm

Spivey et al. (2005) initially applied this mouse-tracking paradigm in language processing research, specifically to explore how phonological competitors influence listeners' real-time processing of spoken words. In this experiment, participants saw a pair of visual stimuli (a target and a distractor image) appearing in the top-left and top-right corners while listening to a pre-recorded audio instruction (e.g., "Click on the candle"). Different types of paired visual stimuli are presented in two conditions (cohort and control condition). For example, participants saw images of a *candle* and a *candy* in the cohort condition, where two images correspond to two words sharing phonological overlap, but a *candle* and a *jacket* in a control condition. Spivey et al. found that listeners showed a curvature toward the distractor (e.g., candy) before eventually clicking on the target image (e.g., candle). However, they failed to observe this curvature in the control condition where the distractor was not representing a word whose initial phonemes overlapped with the target word.

Using a mouse-tracking paradigm, Spivey et al.'s findings support a processing account in which listeners integrate acoustic-phonetic input gradually over time during spoken-word recognition. Their findings correspond to results shown in eye-tracking studies (e.g., Magnuson, Tanenhaus, Aslin,& Dahan, 2003; Spivey & Marian, 1999).

More importantly for our purposes, their study motivated the use of mouse-tracking for exploring real-time language processing.

When investigating listeners' real-time processing of the scalar quantifier *some*, Tomlinson Jr., Bailey and Bott (2013) used a simillar mouse-tracking paradigm with a sentence verification task, as shown in Figure 4.2.

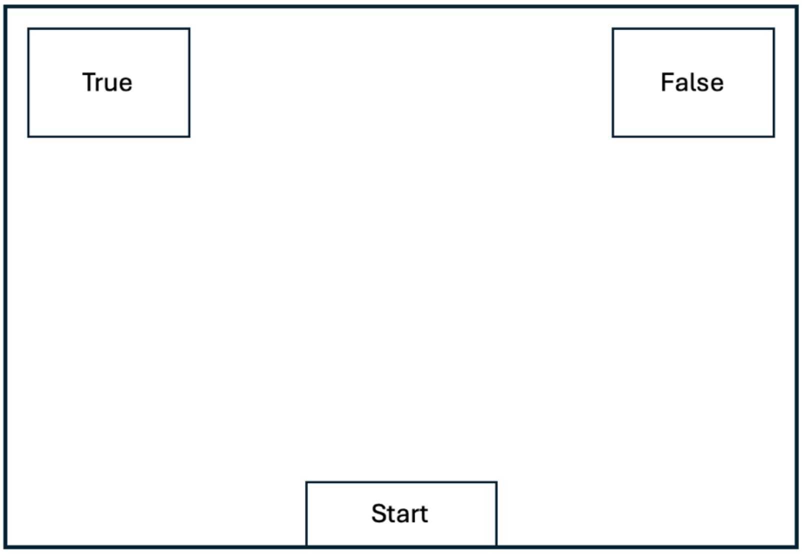


Figure 4.2 A visual display example of the mouse-tracking paradigm used in Tomlinson Jr., Bailey and Bott (2013).

They specifically tested whether the scalar implicature interpretation of *some* ("some but not all") is generated automatically in one step or requires a derivation in two steps where a basic meaning is accessed first and then enriched. Participants were asked to listen to sentences like "Some elephants are mammals" and to verify whether the sentences were True or False. A response to click on True means a participant interprets *some* literally as "some and possibly all", while clicking on False corresponds to the scalar implicature interpretation "some but not all". Mouse-tracking data revealed that when interpreting sentences like "Some elephants are mammals", participants' mouse showed initial movements towards the True response (literal interpretation) but changed direction mid-flight to select the False response (scalar implicature). This pattern of mouse movements captures a competition between interpretations in real time. Results from Tomlinson Jr. et al. (2013) add crucial evidence for a stepwise derivation of scalar implicature, which corresponds to findings in previous studies using an eye-tracking paradigm (e.g., Huang & Snedeker, 2009).

Also to examine the real-time comprehension of the scalar quantifier *some*, Loy, Rohde and Corley (2019)[10] combined a visual world paradigm with mouse-tracking, as suggested in Spivey et al. (2005). Loy et al. (2019) tested whether a speaker's manner of delivery (i.e., whether a speaker is disfluent or not) influences listeners' interpretations of *some* in a given social context. They asked participants to view a visual display as in Figure 4.3 while listening to recorded utterances like "I ate *[uh] some* crackers".

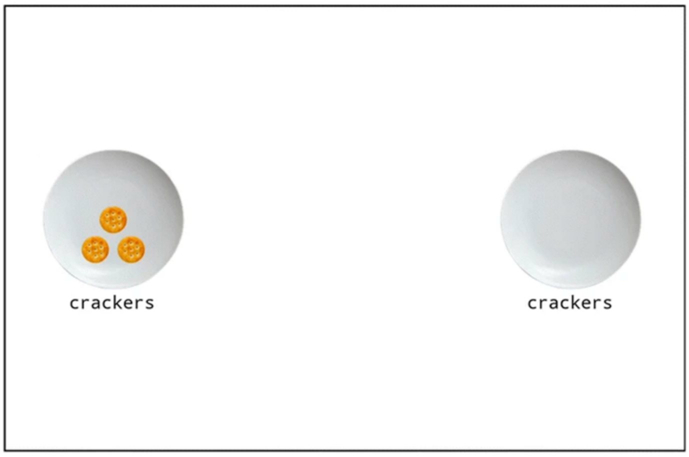


Figure 4.3 A visual display example of the mouse-tracking paradigm used in Loy, Rohde, and Corley (2019).

Loy et al. provided participants with a social context exploiting the concept of face (Goffman, 1967), where speakers in the recorded audios described the number of snacks they had eaten. The more snacks a speaker had eaten (therefore the fewer snacks left in the plate), the greedier the speaker might seem to be. With this given social context, together with the manipulation of speaker's disfluency, participants were expected to click on one of the given two plates to indicate how many snacks they believed the speaker had eaten. In the critical trials, the left plate always represents a pragmatic interpretation (corresponding to meaning "some but not all") and the right plate is an empty plate (corresponding to literal meaning "some and possibly all"). The empty plate indicates that the speaker had eaten all of the snacks and was taken to represent the socially undesirable interpretation of *some*.

The visual display in Loy et al. (2019) showed the starting mouse position, as well as two images along one horizontal axis, unlike the studies outlined above.

[10] This study is covered in detail in Chapter 5.

Despite this change of the visual display, both eye-tracking and mouse-tracking data revealed a pattern whereby, following disfluent utterances ("I ate, *uh, some* snacks"), participants were more likely to interpret *some* literally ("I ate some and possibly all snacks") compared to following fluent utterances. Moreover, Loy et al. showed that differences between conditions emerged quickly after *some* onset with trajectory analysis. Loy et al. not only explores the online effect of disfluency on listeners' processing of some, but also demonstrates the applicability of a combined use of eye-tracking and mouse-tracking methods.

Studies like Spivey et al. (2005) and Tomlinson Jr. et al. (2013) that have used the mouse-tracking paradigm replicated results from previous studies using eye-tracking paradigms. Moreover, Loy et al. (2019) showed an example study of combining the visual world paradigm and mouse-tracking to test a proposed account of real-time processing (also see King, Loy, & Corley, 2017; Loy, Rohde, & Corley, 2017). Together, these studies open up possibilities of using mouse-tracking indepedently to explore real-time language processing. Importantly, Loy et al. (2019) showed that similar conclusions could be drawn from eye- and mouse-tracking, further validating the use of mouse-tracking in this thesis.

4.2.2 Analysing mouse-tracking data

There has been little standardisation for reporting and analysing mouse-tracking data. However, as mouse-tracking and eye-tracking data share the feature of providing trajectory information, methods for analysing eye-tracking trajectories can also be applied to mouse-tracking data. In this section, I do not intend to exhaustively review every method of analysing eye-/mouse-tracking data[11], but to introduce the most common analysis methods used in analysing trajectories of eye-/mouse-movements. These methods include Area Under Trajectory Curve (AUC), Generalized Additive Modelling (GAMM) and Bootstrapped Differences of Timeseries (BDOTS). The overall goal of these methods is to measure processing difficulty. In the following subsections, I briefly introduce features of mouse-tracking data before describing analysis methods. I then introduce how each of the three mentioned analysis method

[11] Useful reviews of analysis methods of eye/mouse-tracking data can be found in, for example, Ito & Knoeferle (2022); Maldonado, Dunbar, & Chemla (2019); Schoemann, O'Hora, Dale, & Scherbaum (2020), etc..

works and their advantages and disadvantages in analysing temporal data with examples from previous studies. I then conclude that BDOTS is the most appropriate method for the current mouse-tracking studies.

Mouse-tracking data

Mouse-tracking data is a time-series of [x,y] coordinates that records the position of the mouse cursor over time as participants engage in a task. Each coordinate pair corresponds to the mouse's position at a position at a specific time point, creating a continuous stream of data points that reflect the trajectory of the cursor.

When analysing mouse-tracking data, autocorrelation is a significant characteristic of mouse-tracking data that needs to be accounted for. Autocorrelation in mouse-tracking data refers to the phenomenon where the position of the mouse cursor at any given time point is correlated with its positions at preceding and subsequent time points. This correlation arises as mouse movements are generally smooth and continuous, meaning that the cursor's location at time *t* is not independent of its location at time *t-1*, *t-2*, and so forth. This temporal dependency makes it important to address autocorrelation in analysing mouse-tracking data, to capture the nature of this kind of time-series data and to make valid conclusions about the underlying cognitive processes. When reviewing the common analysis methods in the following subsections, whether autocorrelation is accounted for is one important evaluating factor.

Analysis Methods

Area Under Curve (AUC)

AUC represents the geometric area between the actual trajectory and a hypothetical straight path from the starting point to the ending point as shown in Figure 4.4 (Freeman & Ambady, 2010). AUC is straightforward to calculate and interpret, as it provides a clear numerical value representing the degree of deviation of the actual trajectory to the ideal straight path. The further a trajectory is away from the idealised straight path, the larger AUC it has. Therefore, in a typical mouse-tracking paradigm, AUC is able to reveal how large the deviation towards a distractor image is even when the target image is eventually chosen. In this sense, AUC is suitable to describe this deviation and generally represent how difficult the processing is (Freeman & Ambady, 2010).

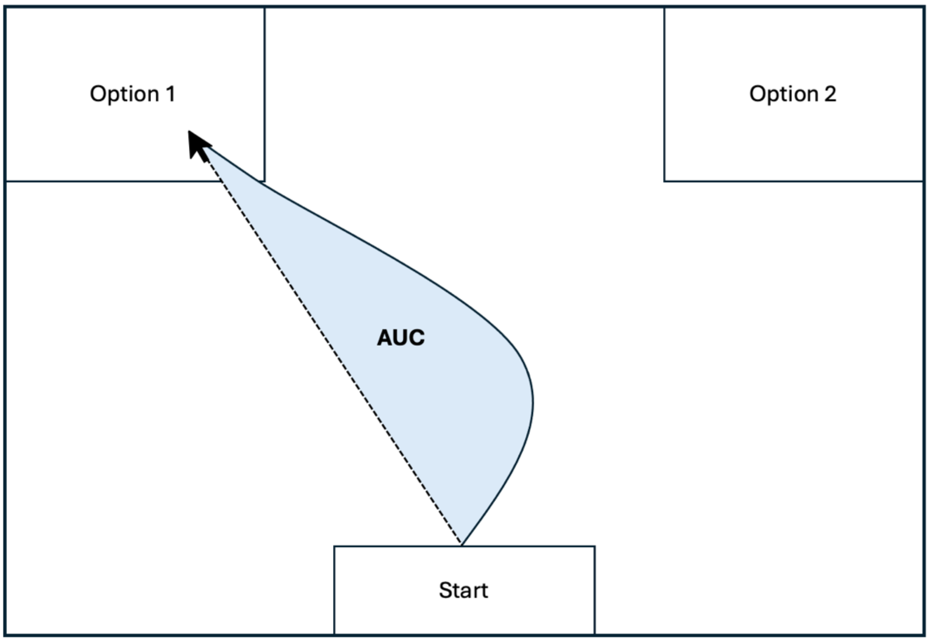


Figure 4.4 A visual representation of Area Under Curve (AUC) in a basic two-item mouse-tracking paradigm, with the blue-shaded area representing the AUC of the shown mouse-trajectory.

Tomlinson Jr. et al. (2013) used AUC to measure participants' mouse movements towards response options (True/False) when verifying sentences like "Some elephants are mammals". Participants who were biased towards a literal interpretation (clicking True at the end) showed smaller AUC values indicating more direct movements towards the True response. But participants who were biased towards a pragmatic scalar implicature interpretation (clicking False eventually) showed larger AUC values, indicating greater deviations towards the competitor response True before settling on False. Combining visualising mouse trajectory and AUC analysis, Tomlinson Jr. et al. presented evidence for a two-step account of scalar implicature derivation.

However, despite the advantages of mapping the deviation into numerical values, AUC fails to capture the temporal features of processing, specifically *when* the deviations occur during the real-time process. This limitation arises because AUC provides a single summary value for the entire trajectory, overlooking the nuances of when specific deviations occur. Also, the way AUC is calculated determines its sensitivity to outliers. Due to this characteristic, AUC is considered better in indexing the overall attraction toward the unselected option (Freeman & Ambady, 2010). However, this sensitivity also means that a large deviation in any point of a trajectory can disproportionately affect AUC, which potentially masks real features of the online processing.

Generalized Additive Modelling (GAMM)

GAMM as an extension of Generalized Linear Models (GLMs), allows for modelling non-linear relationships. GAMM can model the complex non-linear time-course eye-/mouse-tracking data and importantly it can capture when an effect of interest takes place (Wieling, 2018; Wood, 2011). GAMM uses thin plate regression splines to fit curves. A thin plate regression spline is a specific type of spline used in statistical modelling to fit smooth curves to data. Compared to polynomials, a critical advantage of using this type of spline is avoiding overfitting or underfitting problems. Unlike fixed-degree polynomials (linear, quadratic, cubic or etc.), thin plate regression splines adapt their complexity to the data rather than being restricted to a specific degree of flexibility. Moreover, GAMM reduces false positives by accounting for autocorrelation through an AR1 parameter, which adjusts the confidence levels based on the correlation between adjacent time points (Ito & Knoeferle, 2022).

King, Loy, Rohde and Corley (2020) used GAMMs to analyse the eye- and mouse-tracking data when investigating how listeners use nonverbal cues to judge deception during real-time communication. Their study aims to test if and when listeners incorporate a speaker's nonverbal behaviours, such as trunk movements and adaptor gestures like fidgeting, into their on-line judgements of veracity. Participants' eye- and mouse-movements were tracked when watching videos of speakers potentially lying about the location of hidden treasure and choosing between a target object and a distractor object. GAMM captures the dynamic nature and accounts for the temporally correlated nature of eye- and mouse-movements. It also helps identify when participants begin to respond to nonverbal cues in the time course of their processing. The GAMM analysis revealed non-linear differences in eye fixation patterns over time between conditions with and without nonverbal cues, showing a shift in fixation patterns when adaptor gestures are present. Participants' mouse movements were also found to be biased toward a distractor object more often when non-verbal cues were present. With the use of GAMM, King et al. were able to visualise and reveal listeners' integration of a speaker's nonverbal cues into real-time judgement of deception.

Despite its capabilities, GAMM has certain limitations. One limitation is that although GAMM estimates the onset and duration of effects, it does not provide information regarding whether the onset of the effect is significantly earlier or later in one condition than the other. The other is that when a dependent variable is binomial, GAMM cannot account for autocorrelation issues, and data transformation is required to solve this problem. Also, conducting GAMM analysis can be computationally intensive, especially with large datasets, and due to its complexity of implementation and interpretation, GAMM is more suitable for researchers with a high level of statistical expertise.

Bootstrapped Differences of Timeseries (BDOTS)

BDOTS is a non-parametric statistical approach that can identify significant differences between conditions or groups in time-series data (See Oleson, Cavanaugh, McMurray, & Brown, 2017 and Seedorff, Oleson, & McMurray, 2018 for details). BDOTS does not assume a specific distribution of the data, making it robust for different data types. BDOTS involves repeatedly resampling the data (therefore bootstrapping) to create distributions of possible outcomes. For each resample, the difference between the time-series data in two conditions is calculated. By comparing the observed differences to the bootstrapped distribution, BDOTS can estimate whether the differences are statistically significant.

Critically, BDOTS can detect significant differences at specific time points with both analysis results and visualisations, providing a detailed picture of when time-series data in two conditions start to significantly differ from each other. This feature makes BDOTS an appropriate method when the research question concerns the onset of an effect and for how long an effect lasts. Moreover, as an analysis method suitable for highly correlated temporal data, BDOTS also include AR1 (as discussed in GAMM) to handle the autocorrelation issue.

BDOTS has been implemented in R as *bdots* package (Nolte, Seedorff, Oleson, Brown, Cavanaugh, & McMurray, 2023) with vignette introducing detailed steps of analysis. At the time of writing there are two fitting functions available in this package, four-parameter logistic and double-Gaussian, which is often suitable for visual world paradigm data. However, the package also allows additional curve functions to be defined for other types of time-series data (Seedorff et al., 2018).

While BDOTS is a powerful method for detecting time period of significant differences between conditions, conducting BDOTS comes with several challenges. The first challenge is when fitting curves, it is possible that a specific function fits well for some participants' trajectories but not for others', which may affect the statistical results. This potential problem therefore requires a detailed check of goodness of fit when conducting BDOTS analysis. BDOTS provides both visualisations of curve-fitting and fit codes using R^2 as an estimate of goodness of fits.

The second challenge is that large numbers of observations in each trial and large number of participants are required for accurate analyses. In an analysis of eyetacking data, Ito and Knoeferle (2022) used only 16 critical trials as an example and they failed to fit good curves to their data or to replicate their results they had obtained using other analysis methods like Growth Curve Analysis. They emphasised the importance of obtaining sufficient numbers of observations for a successful BDOTS analysis by listing the number of critical trials in previous studies that successfully implemented BDOTS methods (e.g., 384 trials used in Hendrickson, Apfelbaum, Goodwin, Blomquist, Klein, & McMurray, 2021; 480 trials in Sarrett, Shea, & McMurray, 2022).

To conclude, BDOTS approach benefits the analysis of time-series data like eye- or mouse-tracking data in that it is able to detect the specific time period when an effect of interest is significant. Compared to methods like AUC, BDOTS allows for a precise look of temporal change of trajectories in each time point and estimates the significance of differences in two conditions. Also, BDOTS allows to use a specific function (four-parameter logistic, double Gaussian or user-specific functions) to improve the curve fit.

4.3 Current mouse-tracking studies

The current thesis presents four mouse-tracking experiments to explore whether and how a specific type of disfluency, filled pause *uh*, influences listeners' real-time interpretations of the ambiguous word *some*. As previously mentioned, Tomlinson Jr. et al. (2013) using mouse-tracking and Loy et al. (2019) combing the use of eye- and mouse-tracking, have explored listeners' online processing of *some*.

Both studies used two-item paradigms, effectively providing two interpretations of *some* in each critical trial and asking the listener to select their preferred interpretation.

However, as a scalar quantifier, interpretations of *some* should be considered to be on a scale of meanings and not necessarily limited to only two interpretations. Therefore, building on the previous basic paradigm, I designed a four-item paradigm for the series of mouse-tracking studies in this thesis. This paradigm broadens listeners' options of potential interpretations of *some* to four and is thus expected to provide a more natural and more subtle picture of listeners' real-time processing.

Detailed descriptions of the mouse-tracking paradigms used in each study are presented in Chapters 5 and 6. In the following sections, I describe generally what a four-item mouse-tracking paradigm in this thesis is like. I then generalise the features of mouse-tracking data I reported in each mouse-tracking study and the analysis method used in my mouse-tracking studies.

4.3.1 Current paradigms

All mouse-tracking studies in this thesis involved a web-based clicking task using a four-item mouse-tracking paradigm. In this task, participants were presented with four items while listening to audio recordings. An example of a four-item paradigm is shown in Figure 4.5.

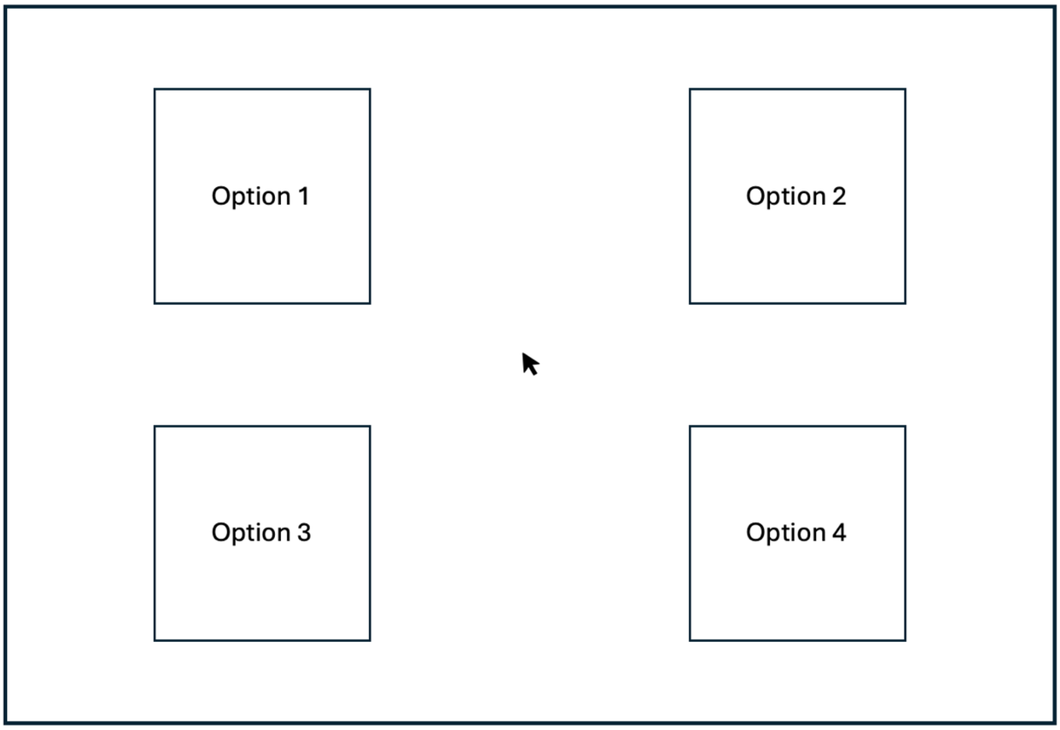


Figure 4.5 A visual display example of a four-item mouse-tracking paradigm

This paradigm provides four choices that appear in the centre of each quarter of screen. The starting position for the participants' mouse was controlled to ensure that the mouse movements in each trial start from the centre of the screen and end at the point when the participant clicks their selected image.

4.3.2 Current analysis methods

I reported and analysed the following features of mouse-tracking data in each study: the **outcome** of interpretation (i.e., final clicks) and the **process** of interpretation (i.e., mouse trajectories). It should be noted that the details for the data visualisations and analyses were slightly different according to specific data in each study (see Results sections in Chapters 5 and 6), but I present here the common information I reported for all mouse-tracking data.

Final clicks

To examine the outcome of participants' interpretations, final clicks (i.e., which image a participant clicked eventually) were recorded. Two perspectives regarding final clicks were reported: clicking results (i.e., the image clicked) and response time (i.e., time between the word *some* onset and a click on the chosen image). Variables of interest (clicking results and response time) were also modelled using mixed-effects regression (logistic or linear) to test whether any observed differences were statistically significant.

Mouse trajectories

Mouse trajectories were analysed to observe participants' processing. For each trial for each individual, mouse trajectories were time-normalised into 101 time-steps for further analysis (following Dale, Kehoe, & Spivey, 2007 and Spivey et al., 2005). This normalisation process allows us to compare trajectories across different trials and participants, despite variations in the absolute duration of each trial. Normalising time in this manner also has important implications for interpreting findings regarding the timing of the comprehension process. Specifically, it allows us to focus on the relative timing of effects within the trial. This process-related view of time enables us to address questions such as whether an effect occur early or late within the trial.

Among the previously introduced common methods of analysing trajectories, the Bootstrapped Differences of Timeseries (BDOTS) was chosen. There are two main reasons for this choice. First, BDOTS provides a means of detecting the specific time period when trajectories start to significantly differ between two conditions. This feature allows us to examine the temporal dynamic of participants' mouse movements. Mapping this significant time period of differences by conditions provides us a clearer picture of how participants' mouse movements reflect the onset of disfluency. The second reason to choose BDOTS is that when the dataset is large, it is relatively more easy to implement and more computationally tractable compared to other similarly capable methods like GAMM.

The current mouse-tracking studies took the following actions to overcome the potential challenges when implementing BDOTS analysis. To obtain sufficient observations in critical trials, a minimum of 150 participants were recruited for each mouse-tracking experiment. Also, to ensure the goodness of fit, both the fit codes and visualisations of curve-fitting were checked in detail. Moreover, to aid the interpretation of BDOTS analysis results, I used a 'rainbow' visualisation of participants' mouse trajectories by final clicks and by conditions (disfluent/fluent) when examining the mouse trajectories. Figure 4.6 shows an example of one such visualisation in Chapter 6.

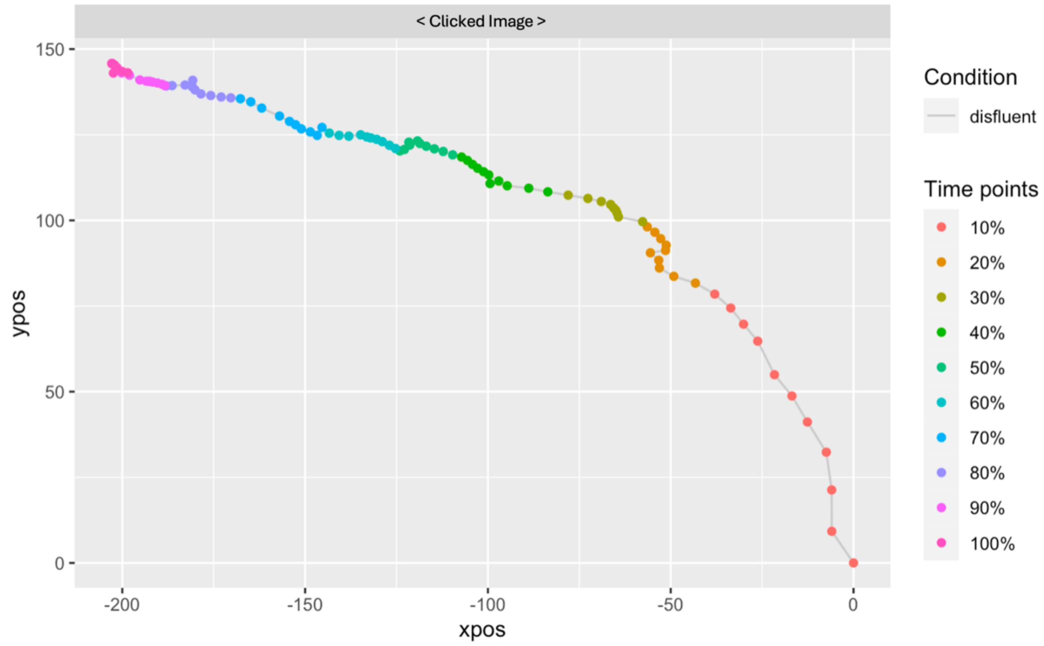


Figure 4.6 An example of visualising mouse trajectory in a four-item mouse-tracking paradigm (one quadrant from a larger visualisation in Experiment 3, Chapter 6).

Figure 4.6 shows an average literal mouse pointer path starting from the centre of the screen (bottom right in the current figure) and in this case ending on the left top of the screen where the clicked image is located. This line, therefore, shows what the average mouse movement looks like for participants who clicked on this specific image in one particular condition. Coloured points were added for time-related characteristics. The colour of points (from red points at the centre of the screen to the violet colour in the corner) indicates successive 10% increments of trial time. Together with BDOTS analysis results, this way of presenting mouse trajectory information offers a robust examination of participants' process of comprehension.

4.4 Conclusion: Mouse-tracking in understanding language comprehension

Mouse-tracking, like eye-tracking, is an effective method for exploring the real-time unfolding of listeners' processing in language comprehension. Using a typical two-item paradigm, previous studies have demonstrated the utility of mouse-tracking methods for investigating the temporal dynamics of language processing, and have also established the possibility and benefit of using mouse-tracking method to collect large-scale web-based data. Using an appropriately adapted four-item paradigm, the mouse-tracking studies in this thesis enable a deeper understanding of listeners' real-time processing of the scalar quantifier *some*. To my knowledge, these studies are the first to use four items in a visual mouse-tracking paradigm. By developing this new paradigm specifically for the current research question and providing standardised ways of reporting both outcomes and processes of mouse-tracking tasks, this thesis affirms the potential of using mouse-tracking techniques in studying real-time language comprehension.

Chapter 5
When *some* is more: Disfluency in context

5.1 Introduction

When it comes to language comprehension, the content of utterances is not the only important factor. The speaker's manner of delivery may also influence the interpretation of the utterances' meanings. Spontaneous speech produced with disfluencies, like *uh* or *um*, is thought to be less certain (Smith & Clark, 1993) or less confident (Brennan & Williams, 1995). Such disfluent speech is also found to influence listeners' judgement towards whether an utterance is truth or a lie (Loy, Rohde, & Corley, 2017; Zuckerman, Koestner, & Driver, 1981). These findings highlight the relationship between a speaker's manner of speech and listeners' pragmatic interpretations of disfluent utterances. However, there still remains the question of how this connection between disfluencies and listeners' pragmatic interpretation is formed: whether listeners' interpretation is affected by disfluencies via *speaker modelling*, or via a *heuristic* association of disfluency and uncertainty or lying.

For example, if listeners' interpretations of disfluencies are based on heuristics, they might automatically associate fillers to uncertainty or deception, regardless of the speaker's actual confidence or truthfulness or contexts. However, if listener interpret disfluencies via modelling speaker's perspectives, they might infer from disfluencies that the speaker is unsure or lying about the utterances in the given context.

Critically, disfluencies have different effects in different contexts. The current study aims to explore how disfluency affects listeners' comprehension given a context, addressing a main question of this thesis (heuristic or speaker modelling) by focusing on the ambiguous word *some* in a social context.

Social contexts were found to influence how people perceive disfluencies in sentences with ambiguous word *some*, and several studies have exploited the ambiguity between the literal ("some and possibly all") and the pragmatically strengthened ("some but not all") interpretations of *some* (Bonnefon, Dahl, & Holtgraves, 2015; Bonnefon, Feeney, & Villejoubert, 2009; Loy, Rohde, & Corley, 2019). Consider a sentence like "*Some* people like your idea". Previous studies exploiting the concept of face found that interpretations of *some* were influenced by whether face-threatening content was involved (Bonnefon et al., 2009). Bonnefon et al. (2015) investigated how disfluencies, specifically silent pauses, influenced the interpretations of *some* in such contexts where the speaker's politeness goals were involved. They asked participants to read vignettes like (1) and answer the question of to what extent participants thought the sentence to guarantee a literal interpretation (i.e. "Possibly all people hated/loved your idea"). Whether the speaker was described to stay silent before delivering a face-threatening or face-boosting expressions was manipulated.

(1) Yesterday, you pitched an idea to a group of five persons. Today, you ask Bob (who was in the group) what people thought of your idea. Bob <*stays silent for a few seconds. Then he*> replies: "*Some* people *hated/loved* your idea."

Bonnefon et al.'s results showed that when the description of the speaker remaining silent was present, participants always preferred an unpleasant interpretation of *some*. Specifically, in a scenario where face-threatening message was delivered (i.e., *Some* people hated your idea), people tended to rate higher for a literal interpretation ("Possibly all hated your idea"). However, in the face-boosting scenario (i.e., *Some* people loved your idea), people still preferred an unpleasant interpretation, but in this case, a pragmatic interpretation ("Not all people loved your idea"). This study provided off-line evidence for disfluency's contextual effect on people's interpretations of the ambiguous word *some*.

Also exploiting the concept of face, Loy et al. (2019) used a visual world paradigm study and provided further real-time evidence for the effect of disfluency on the interpretation of *some*. Loy et al.'s study used a social context involving potential face-loss by admitting one's greediness. They described a cover story: participants were given snacks to eat freely while watching a documentary film. The true purpose of the study, investigating greed, was only revealed afterwards. Participants needed to report the number of snacks they consumed. Participants in Loy et al.'s studies were told to listen to the recorded answers from those taking part in the greed study and choose from two plates with a number of **remaining** snacks to indicate how many snacks they believed the speaker to have actually eaten (i.e., one snack left on the plate meaning six snacks eaten). In each critical trial, the screen display (left panel in Figure 5.1) showed one plate corresponding to a pragmatic interpretation of *some* (e.g., ate some but not all snacks), and an empty plate compatible with literal interpretation (e.g., ate possibly all snacks). Speakers' manners of delivery were manipulated for critical trials, which were of the form "I ate some oreos" (or other snack) in the fluent condition and "I ate, uh, some oreos" in the disfluent condition.

Both eye- and mouse-movements showed that where speakers were disfluent ("I ate, *uh, some* oreos"), listeners were more likely to make an early commitment to a literal interpretation ("I ate possibly all oreos") than when the speakers were fluent. These results suggested that listeners took not only the social context, but also the speaker's manner of delivery rapidly into account while processing the meaning of *some*.

Loy et al. also included a control display (right panel in Figure 5.1) to further explore how listeners' interpretations of *some* was affected by disfluencies. In the control display, the empty plate from critical display, compatible with the literal meaning of *some*, was replaced by a plate with only one snack eaten. They considered this plate incompatible with the meaning of *some*, and therefore listeners were expected to only choose it if they thought the speaker was lying. Evidence from eye- and mouse-movements indicated that for control trials, participants in fact showed early bias towards the only plate compatible with any meaning of *some* instead of the plate with only one snack eaten. Loy et al. concluded that this result was consistent with the view that the effect of disfluency in the given context was better explained by potential face-loss concerns rather than deception.

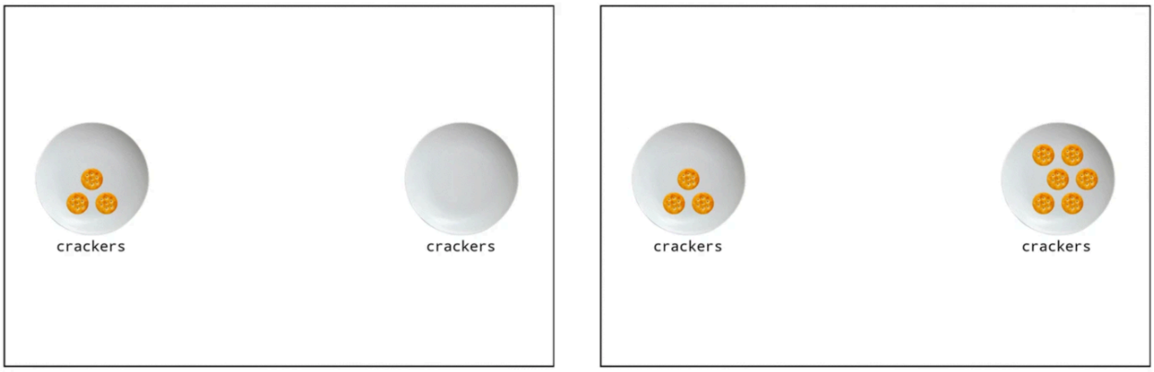


Figure 5.1 Example of screen displays used in Loy et al. (2019). *(Left Panel)* Critical displays: the left plate depicting the pragmatic interpretation of *some* and the right plate depicting the literal interpretation; *(Right Panel)* Control displays: the left plate depicting the pragmatic interpretation of *some* and the right plate with the meaning of "one".

Loy et al.'s study provided two important conclusions. First, listeners were sensitive to social cues hidden behind the speaker's manner of delivery, as they showed a rapid integration of these cues to real-time language processing. Second, methodologically, the corresponding results from eye- and mouse-movements in this study demonstrated that mouse-tracking could be successfully used in the context of the visual world paradigm.

Despite these indications, one question that remained unclear from Loy et al. is how disfluency exerts its effect in a given social context. Setting up the control trials to investigate whether disfluency cues lying or face-saving under the given social context might be an attempt to touch upon this question. However, their results could not be the whole story. A critical reason is that Loy et al. did not vary their social contexts and therefore their study failed to tell apart whether the preferred literal interpretation of *some* in their given context was due to listeners' reasoning cued by disfluency, or merely a heuristic association of the literal interpretation and the given context. The second reason is that a plate with one snack eaten in the control display might not be the best display representing the account that disfluency signals deception. By definition, the literal interpretation of *some* includes the lower bound meaning of "at least one". Therefore, "one" can be seen as compatible with the meaning of *some*. This controversy might challenge Loy et al.'s conclusion that listeners associate disfluencies with face-loss rather than lying. The third reason is regarding the complexity of the task itself: with the screen display designed as plates with a remaining number of snacks, participants had to not only process the utterances, but also do a calculation using the known total number of snacks and the displayed number of snacks to get the number of eaten snacks and then make the decision of which plate to choose. Even though it was a simple calculation, the cognitive effort cost in this indirect display of represented number might influence the results in eye- and mouse-movements.

5.1.1 Present study

In this chapter, we present a conceptual representation of Loy et al. (2019). There were two things from Loy et al. we kept in the current studies to explore further. First, I kept the main research question of investigating the disfluency's pragmatic role in listeners' real-time processing of the scalar quantifier *some*. Second, I kept a similar methodology by applying web-based mouse-tracking techniques to explore listeners' real-time processing of *some* in a given social context. While conceptually replicating Loy et al. (2019), the two studies presented in the following sections aim to target the remaining questions of Loy et al. by changing two main things. First, I changed a social context presented to participants to test if the pragmatic role of disfluencies remained (when the context was varied). Second, I changed the task to a more direct one (detailed in Experiment sections) which should be more straightforward to participants and could rule out other processes with costly cognitive effort. In this way, the current study hopes to replicate the findings in Loy et al. with this simpler paradigm.

Specifically, the current study included two web-based mouse-tracking experiments to measure the real-time interpretation of *some*. The context used in these two experiments was a job interview for a delivery driver's position. This social context was set up as a job interview could be a common shared life experience that is easy to understand. More importantly, in job interview scenarios, it was natural to involve a speaker's, in this case, interviewee's, goals of building good images of oneself to get the job. As a result, disfluency might be interpreted as the speaker avoiding face-loss or as a cognitive consequence of the speaker's efforts of lying.

Participants in the current two experiments were asked to listen to a series of recorded interview conversations like that in (2), which manipulated whether the interviewees were disfluent or not when answering the interviewer's questions.

(2) Interviewer: “How many inspections has your vehicle failed?”
Interviewee: “My vehicle failed *some* inspections.” (fluent) *or*
“My vehicle failed, *uh*, *some* inspections.” (disfluent)

In each of two web-based clicking tasks, participants saw four images depicting different numbers of interviewee disqualifications while hearing the interview audios. By disqualifications, we refer to things that are undesirable in the given context of a job interview, such as the number of failed inspections or the times of being late for a delivery. Critically different from the two-image paradigm in Loy et al. (2019), four images representing various possible interpretations of *some* were displayed as options. This design of presenting a scale of meanings of *some* was to combine both critical and control conditions of Loy et al. in one display, taking into the consideration that interpretations of *some* could be a range of possible meanings. In Experiment 1, available interpretations of *some* were represented by one, two, four or five disqualifications, covering the lower-bound meaning of *some* (one: “at least one”), pragmatically strengthened meaning (two/four: “not all”), and upper bound of literal interpretation (five: “possibly all”). As the use of the empty plate in Loy et al., the absolute upper bound meaning *all* was included to test whether it was possible for participants to extend the meaning of *some* to this higher end (“actually all”), given a context supporting this direction of social reasoning. As in the current interview context where more disqualifications were undesirable, an interviewee might use *some* to soften the less favoured fact that “My vehicle actually failed all the inspections”.

Participants’ mouse movements throughout the experiment session were recorded, and we measured, for each trial, participants’ clicks (i.e., which image was eventually clicked) and their mouse trajectories made before each click. In line with Loy et al.’s findings, if listeners interpret *some* via a speaker-modelling action of reasoning socially of disfluency, similar results to Loy et al. (2019) were expected to be observed, with preference towards images compatible with literal “some and possibly all” interpretation following disfluency, despite the use of a different social context.

5.2 Experiment 1

5.2.1 Methods

5.2.1.1 Participants

One hundred and sixty (Female: 61, Male: 97, Non-binary: 2) self-reported native English speakers were recruited via Prolific Academic, and each participant was compensated £2 for a task that was estimated to take 15 minutes (£8/h). An additional 21 participants were tested but their data were not included in our analyses, because these participants (1) failed to finish the whole experiment session or their data were not correctly saved in our server, (2) took more than 40 minutes to finish the whole session (this time limit is based on pre-tests with non-native English speaker participants), (3) failed both attention checks, or (4) completely correctly guessed the aim of this experiment. All participants were between 18 and 35 years old, with normal or corrected-to-normal vision, and with no self-reported hearing difficulties. Every participant provided informed consent by clicking a button on an information page at the beginning of experiment, in accord with the University of Edinburgh Research Ethics Committee approval (reference number: 271-2122/3).

5.2.1.2 Material and experiment design

We conducted a within-subjects experiment, manipulating the presence of disfluency (present vs. absent) in critical trials. Each participant listened to 6 interviews, each comprising 6 exchanges between an interviewer and one of 6 interviewees. Each exchange constituted one trial and was associated with a set of four images (Figure 5.2). All of the interviews comprised exchanges on the same six topics, using the same 6 sets of four images. Within each interview, 2 of the exchanges were critical trials; one each with a fluent or disfluent response from the interviewee. Over the six interviews, each type of qualification served once as a critical disfluent and once as a critical fluent item, so that each participant encountered 12 critical trials (6 disfluent).

Figure 5.2 Display examples for four target trials in Experiment 1 (Top-left panel: interviewer asking “How many times have you been late for a scheduled delivery?” and interviewee answering “I have been late for *[uh] some* of the scheduled deliveries”; Top-right panel: interviewer asking “How many times have you had complaints from customers recently?” and interviewee answering “I’ve had *[uh] some* complaints”; Bottom-left panel: interviewer asking “How many ‘F’s did you get for driving proficiency tests?” and interviewee answering “I got *[uh] some* ‘F’s”; Bottom-right panel: interviewer asking “How many inspections has your vehicle failed over the past few months?” and interviewee answering “My vehicle failed *[uh] some* inspections.”)

5.2.1.3 Visual stimuli

Six sets of images indicating counts of qualifications corresponding to six interview questions were used as potential referents in the experiment. The six questions concerned the following qualifications: Q1: Number of times being unpunctual in scheduled deliveries; Q2: Number of complaints from customers; Q3: Number of the cases damaging customers’ parcels; Q4: Number of failed driving proficiency tests; Q5: Number of times getting lost on delivery trips; Q6: Number of vehicles failing inspections.

In each target trial, participants saw four images with one, two, four, or all (five) disqualifications (represented in red in images) – number of red unpunctual clocks versus green punctual clocks (top-left panel in Figure 5.2), number of red unhappy faces versus green happy faces (top-right panel in Figure 5.2), number of red 'F's versus 'A's (bottom-left panel in Figure 5.2), or number of green ticks and red crosses (bottom-right panel in Figure 5.2). For convenience, these images will be further referred to as one-cross, two-cross, four-cross, and all-cross. In response to each interview question, more red representations in the images represent a higher number of disqualifications (i.e., lower qualifications). Images always appeared in fixed locations (one-cross top left, two-cross top right, four-cross bottom left and all-cross bottom right).

5.2.1.4 Audio stimuli

For every participant, the whole experiment session included six complete interviews. Every complete interview comprised six conversations between the same interviewer and a different interviewee, where two of the six conversations were critical trials and the remaining four were filler trials, used to reduce the chance that participants would notice the experimental design. In total, every participant heard 36 exchanges, with 12 critical and 24 filler trials included.

Recording of audio stimuli was approved by the University of Edinburgh Ethics Committee with the reference number 268-2122/4. Six native British English (three males and three females) recorded the interviewee scripts, and one native American English speaker recorded the interviewer script.

In critical trials, the recorded interviewees' answers included *some* in either the disfluent or fluent condition. Interviewees' answers in filler trials contain no use of the disfluency *uh* or the critical word *some*, but have other forms of disfluency (e.g., silent pauses, prolongation, substitution, etc.) and other expressions of quantifiers[12] (e.g., 'a few', 'more/less than', etc.). For example, the interviewee's answer in a filler trial about

[12] To avoid adding extra processing difficulty for listeners (discussed in section 3.4.1.1), vague quantifiers such as "rarely any", "less than half", or "around two or three" were used in filler trials instead of specific numbers. Despite these efforts, there remains a slight possibility that alternative expressions could influence listeners' processing in critical trials. However, since the focus of the experiments was on the effect of disfluency on listener processing and the filler trials were used across experimental conditions, we do not expect the presence of salient alternatives in filler trials to have a significant impact on the main findings.

the number of the failed driving proficiency test is “I got an ‘F’ for half, actually less than half of them”, using substitution as a disfluency indicator and ‘less than half’ to indicate the amount.

The interviewer and interviewee audio were recorded online via a browser-based recording tool implemented in JavaScript.[13] Step-by-step instructions were provided to speakers on how to use the recording website, with details to pay attention to before and during recording. Speakers were asked to record each sentence at least three times for the experimenters to choose the most natural version of utterances.

The collected recordings were further edited using Audacity (version 2.4.2) in the following way: for fluent utterances, the most natural version was chosen and used; for disfluent utterances, the most natural disfluency *uh* was extracted from one disfluent utterance and was added to the other disfluent utterance (replacing the original *uh*), so that the two critical trials in the same interview had the same disfluency *uh*. For fillers, no editing was applied.

5.2.1.5 Procedure

Participants completed the experiment, implemented in JavaScript using jsPsych (De Leeuw, 2015) online. After reading the welcome and information page, they provided signed consent of participating via clicking ‘I consent’ button. Then they answered five demographic questions before experiment session. The main mouse-tracking experiment session started with a trial in which the audio instructed participants to click on a specific image. This was designed to familiarise participants with the experiment, as well as to test that they could hear the audio, allowing them time to adjust the volume as required. Following the audio check, a brief instruction page reviewed the cover story for the study and explained the experimental procedure.

The flow of the experiment is shown in Figure 5.3. At the start of each trial, participants clicked a central ‘Ready’ button, which ensured that the pointer was at the centre of the screen. After clicking the ‘Ready’ button, four images appeared on the screen, corresponding to four possible interpretations of the interviewee’s answer. After 2000ms, the first part of the audio (interviewer question) was played. During this part of the audio, participants could move the pointer using the mouse, but clicks (on any of the four images or elsewhere) were not registered.

[13] Recordings were made during the Covid-19 pandemic under UK social distancing regulations.

Once the interviewee's response started playing, participants were able to click on any of the four images as soon as they wished. If no click was made within 10s of the interviewee audio offset, the trial timed out and the screen automatically showed the 'Ready' button indicating the start of the next trial.

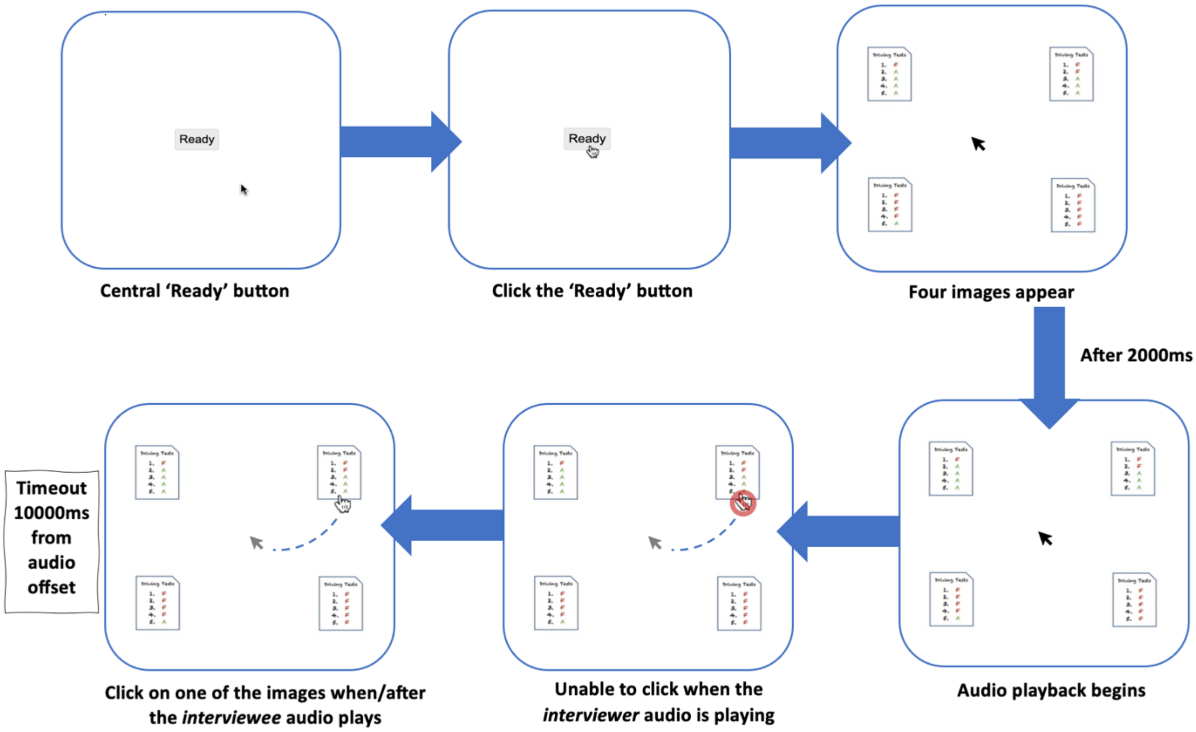


Figure 5.3 Flow of each trial in the mouse-tracking experiment

Two attention-check trials were included in the middle (after the third interview conversation) and at the end (after the last interview) of each participant's session. In these trials, the interviewer gave straightforward information indicating which image should be clicked. For example, in the middle attention check trial, the interviewer said, "Among the candidates we have interviewed so far, I'm thinking about ruling out the one whose vehicle has failed the most inspections over the past few months" such that the interviewer question uniquely determined the answer (left-bottom image in Figure 5.4).

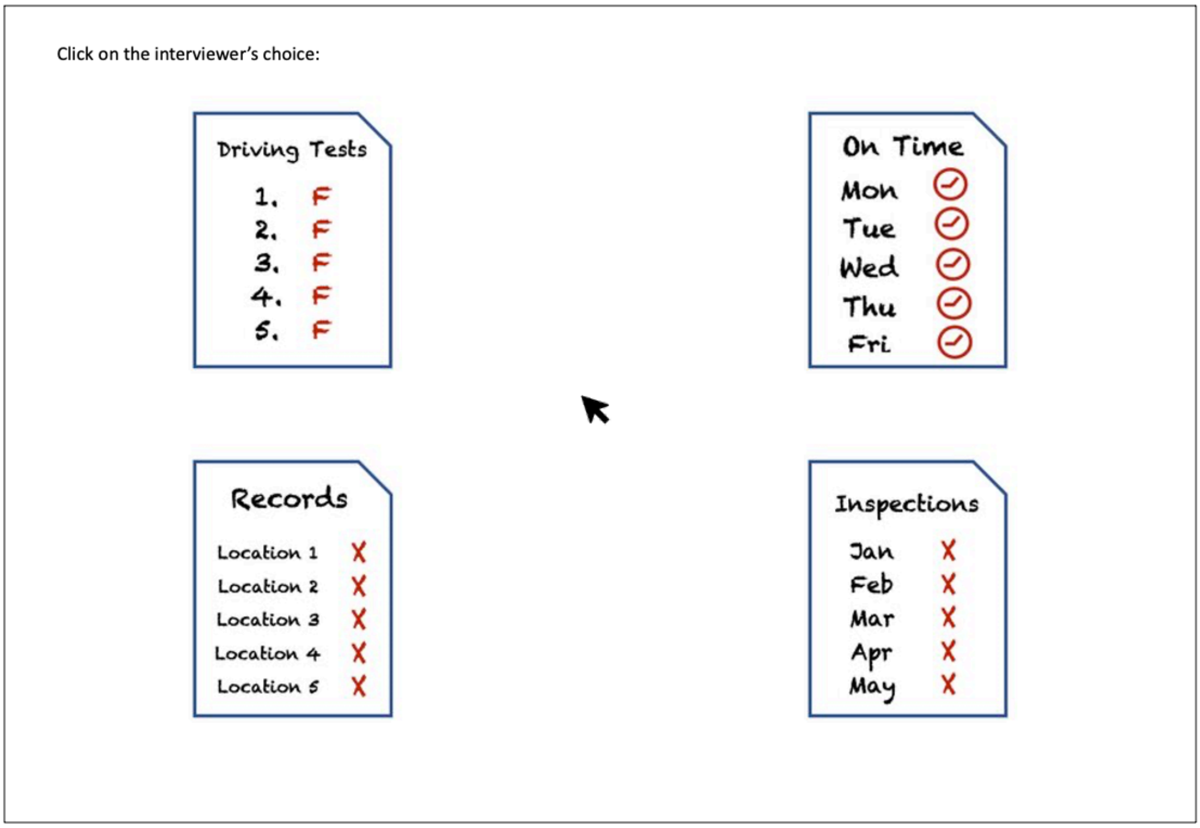


Figure 5.4 Attention-check trial screen display example (mid-attention check)

The mouse-tracking experiment stage finished with the end attention-check trial, after which participants answered two after-experiment survey questions, including a question about suggestions for this experiment and a question of whether they had noticed what the experiment was about.

For each experimental item, we recorded the mouse pointer position over the time as well as the identity of the object clicked.

5.2.2 Results

Data and analysis scripts are available at https://osf.io/m59u3. Prior to analysis, data from filler trials, as well as practice and attention-check trials, were removed. All the statistical analyses were carried out in R version 4.2.3 (R core team, 2023).

5.2.2.1 Final object click

Click results

Total numbers of clicks on each image by condition (disfluent/fluent), as well as the percentages of participants clicking on each image in each condition, are shown in a bar plot presented in the left panel of Figure 5.5.

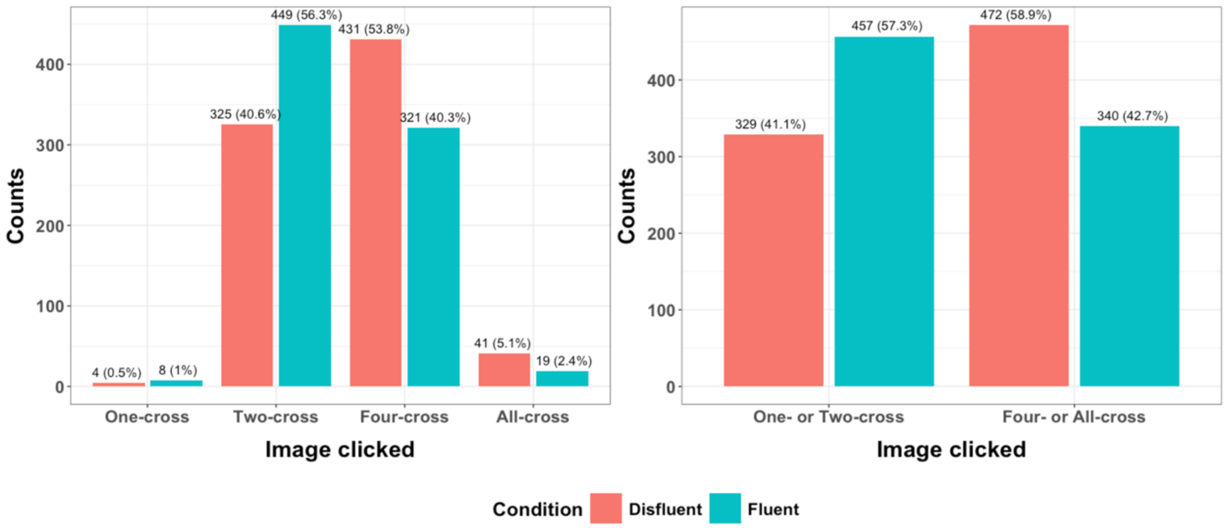


Figure 5.5 *(Left panel)* Bar plot showing total number and distribution of mouse clicks recorded on each image (One-cross, Two-cross, Four-cross, or All-cross) by manner of delivery (disfluent/fluent); *(Right panel)* Bar plot showing total number and distribution of mouse clicks recordesd on each group of choices (images with smaller numbers: One-/Two-cross; with larger numbers: Four-/All-cross) by manner of delivery (disfluent/fluent).

Following disfluent utterances, participants clicked fewer one- and two-cross images, but more four- and all-cross images compared to fluent utterances. For analysis, we separated clicks on these four items into two groups of choices: clicking on images with larger numbers (i.e., four- or all-cross images) and clicking on the ones with smaller numbers (i.e., one- or two-cross images). The right panel in Figure 5.5 shows total numbers of clicks on these two groups of choices (one-/two-cross images vs. four-/all-cross images) by condition (disfluent/fluent). When analysing mouse click results and mouse trajectories in the following steps, we used these two groups of clicking choices when referring to participants' responses. In the following analysis, an effect with p-value less than 0.05 or absolute t value larger than 2 (following Baayen, 2008) is regarded to be significant.

Clicks on the larger-number choice group (clicks on four-/all-cross images) were modelled using a mixed-effects logistic regression. This mixed-effects logistic model includes one fixed effect for the within-participant and within-item predictor, the manner of delivery (fluent vs. disfluent), and random effects with by-participant and by-item random intercepts and slopes. Participants were more likely to click on the larger-number images following a disfluent utterance compared to a fluent utterance (β = 0.990, SE = 0.199, $p < 0.001$). This response pattern of participants fits the predictions of a speaker-modelling action of social reasoning, which suggest that the presence of disfluency in this given context bias interpretation in favour of the socially undesirable meaning, the upper bound of literal "some and possibly all" interpretation of *some*.

Time taken to click

After hearing the critical word *some*, the time participants took to click their final pick of image (grouped into larger- and smaller-number choices) is shown in Figure 5.6. The overall time (after the onset of *some*)[14] spent in deciding which group of image to click was modelled using a linear mixed effects regression. The model included the fixed effects of the manner of speech (fluent vs. disfluent), whether clicking larger-number group of images at the end and their interaction, and random effects with by-participant and by-item random intercepts. There was no significant effect of manner of delivery (β = 16.2, SE = 86.0, t = 0.19) or interactive effect of disfluency and clicking choices (β = -48.6, SE = 124.7, t = -0.39) on participants' time taken to click after hearing *some*. For those eventually clicking on images with smaller numbers (One-/Two-cross images), they took significantly longer time compared to those clicking on images with larger numbers (Four-/All-cross images) (β = 302.8, SE = 96.1, t = 3.15).

[14] We calculated overall reaction time relative to the onset of *some* here to avoid the length feature of different audios influencing the statistical results. As some items' audios were originally longer than the others, therefore participants' reaction time hearing those audios was expected to be longer in nature. Time-locking the overall time to the onset of some was to eliminate this influence.

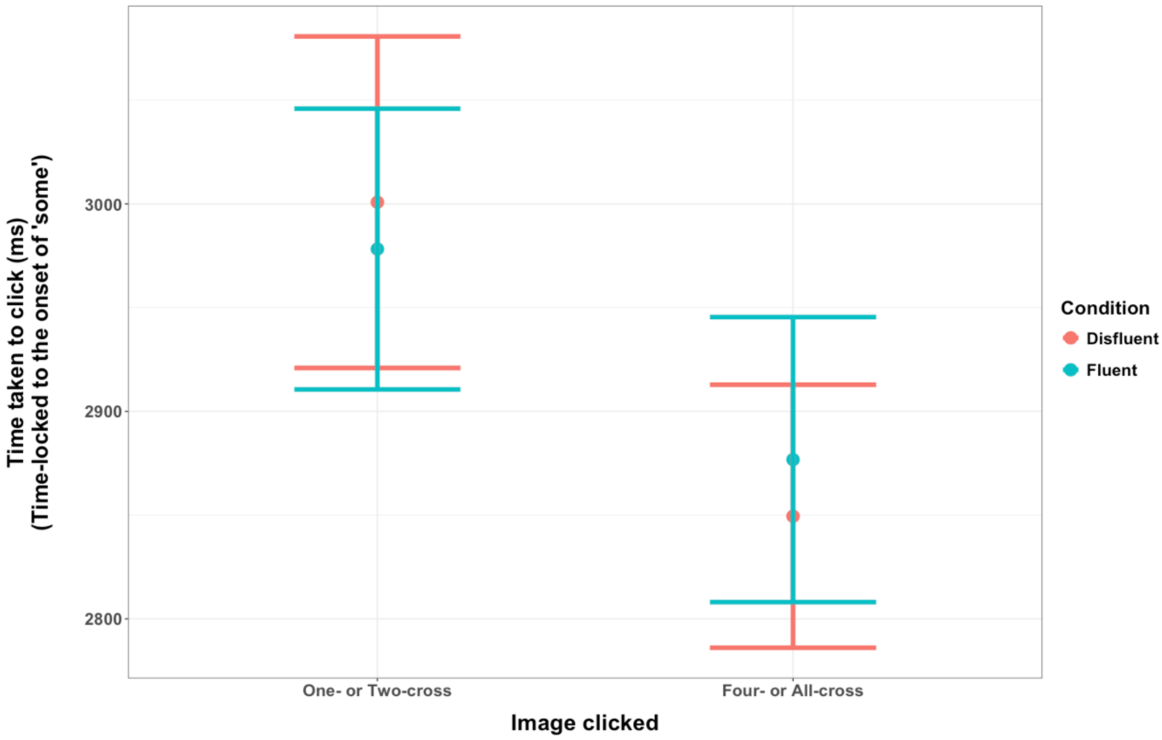


Figure 5.6 Average time to click (relative to the onset of critical word *some*) on each group of choices (images with smaller numbers: One-/Two-cross; with larger numbers: Four-/All-cross) by manner of delivery (disfluent/fluent). Error bars represent ±1 SE of the mean clicking time.

5.2.2.2 Mouse movements

The mouse-tracking data collected from 160 participants were processed prior to conducting the analysis. For each trial for each individual, mouse-trajectories were time-normalised into 101 time-steps (following the time-normalisation introduced in Dale, Kehoe, & Spivey, 2007 and Spivey, Grosjean, & Knoblich, 2005). Step 0 represents the onset of the interviewee audio and step 100 corresponds to the time point when the participant made the click. All movements were calculated relative to an origin (0,0) which represented the centre of the screen.

Participants' mouse trajectories towards each image in each condition (disfluent/fluent) are shown in Figure 5.7. Each line represents the average trajectory of the mouse pointer, starting at the onset of the interviewee's response and ending when an image is clicked. Coloured points indicate relative time. The colours (from red points at the centre of the screen to violet in each corner) indicate each 10% of trajectory time.

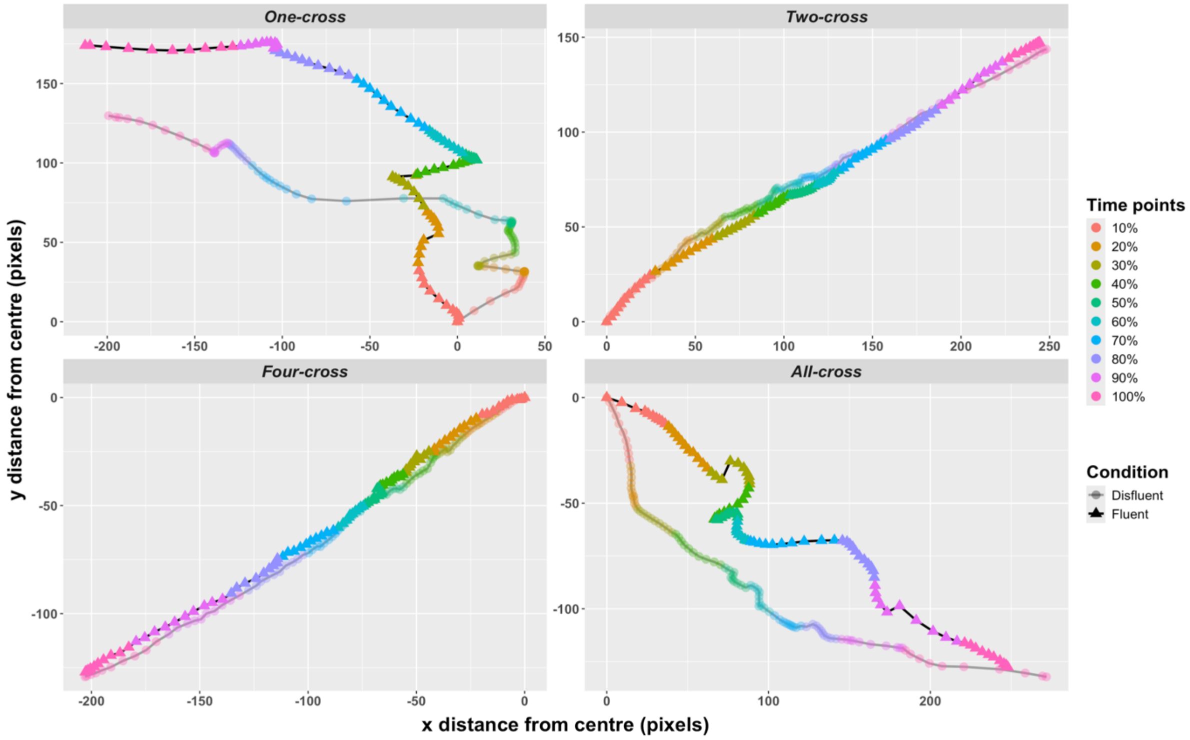


Figure 5.7 Aggregated mouse trajectories towards four images (one-, two-, four-, all-cross) by condition (disfluent/fluent), and the colours of points (from red points at the centre of the screen to violet in each corner) indicate 10%, 20%, 30%...100% of trial time. Note that the disfluent condition is represented with the grey line with round points in relatively lower saturation, and the fluent condition is represented with the black line with triangle points in normal saturation.

From Figure 5.7, we observe that the average mouse trajectories for two- and four-cross images show similar patterns in disfluent and fluent conditions, in terms of both path and distance over time, while the trajectories for one- and all-cross show quite different patterns in each condition. However, we should be aware that the number of clicks on the one- and all-cross images is relatively small: only 0.5% of clicks on one-cross images and 5.1% of clicks on all-cross images in disfluent condition, 1% of clicks on one-cross images and 2.4% of clicks on all-cross images in fluent condition. Given the small numbers of data points for these two images, the following analysis will continue to use the mentioned two groups of choices (larger- and smaller-value images) to explore participants' mouse movements in real time.

5.2.2.3 Mouse trajectory analysis

To explore differences in participants' mouse movement patterns between disfluent and fluent conditions, the Bootstrapped Differences of Timeseries (BDOTS) package (version 1.2.5, Nolte, Seedorff, Oleson, Brown, Cavanaugh, & McMurray, 2023) in R was used to analyse our mouse-tracking data. BDOTS package provides statistical methods designed specifically to handle time-series data, with the advantage of estimating when an effect was significant and how long it lasts (Oleson, Cavanaugh, McMurray, & Brown, 2017; Seedorff, Oleson, & McMurray, 2018). The ability of estimating the time window when two time-curves start to differ makes BDOTS suitable for our real-time mouse-tracking data.

Before performing this analysis, a distance variable representing the distance from each mouse point to the horizontal line in the middle of the screen, referred to as the centre line (see Figure 5.8). This reference line was chosen based on the clicking patterns observed in the final click results (shown in section 5.2.2.1), in which participants showed obvious differences in clicking preference between conditions. There were more clicks on four- and all-cross images (two bottom panels on the screen display as in Figure 5.8) in disfluent condition compared to fluent condition, while more clicks on one- and two-cross images (two top panels on the screen display as in Figure 5.8) were observed following fluent utterances compared to disfluent utterances. Henceforth, we regarded the middle line which divided the screen into halves as a reference line to observe the influence of the condition on participants' mouse movements. Distance to this line, therefore, represents how far a participant move away from a neutral line to a preferred interpretation (either toward larger-numbers or smaller-numbers).

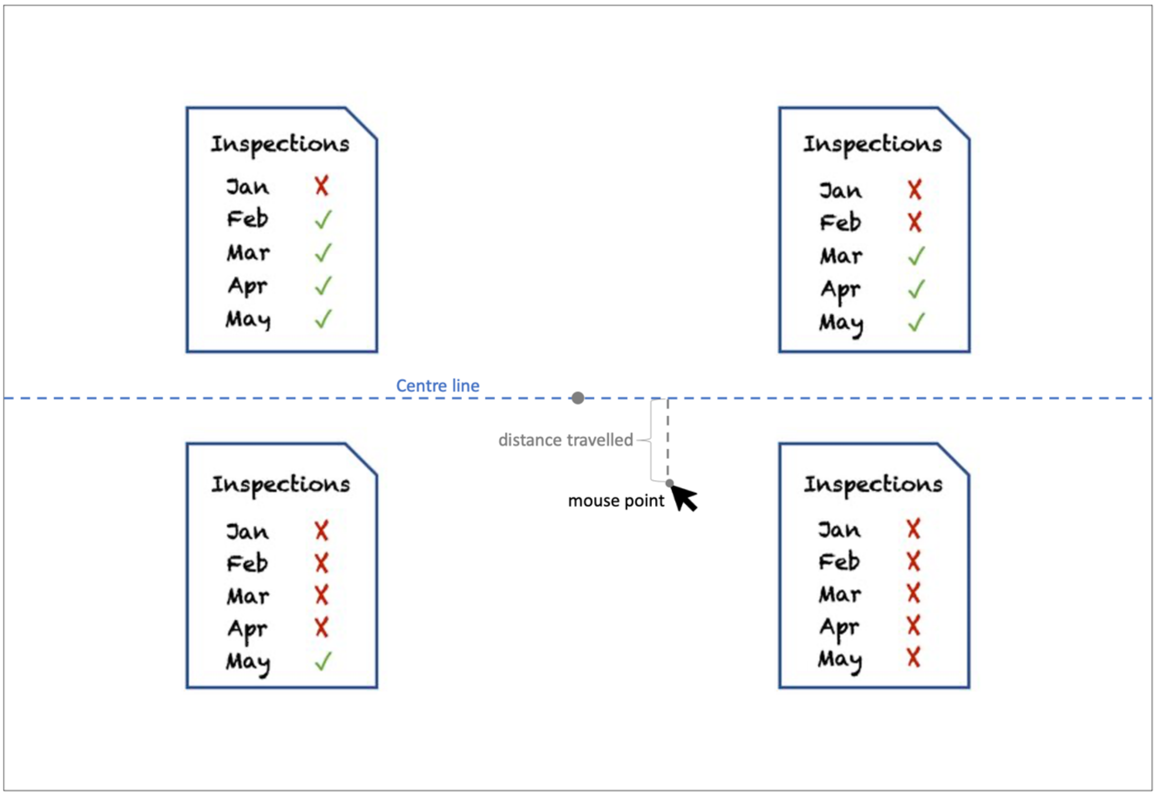


Figure 5.8 A visual representation of the distance from a given mouse point to the horizontal centre line of the screen display.

The proportions of participants' mouse movements towards each group of clicking images (images with larger-/smaller-numbers) across time in each condition (disfluent/fluent) are shown in Figure 5.9. Each line represents the proportion of mouse distance travelled from the horizontal centre line of screen towards the clicked image, starting at the onset of the interviewee's response and ending when an image is clicked.

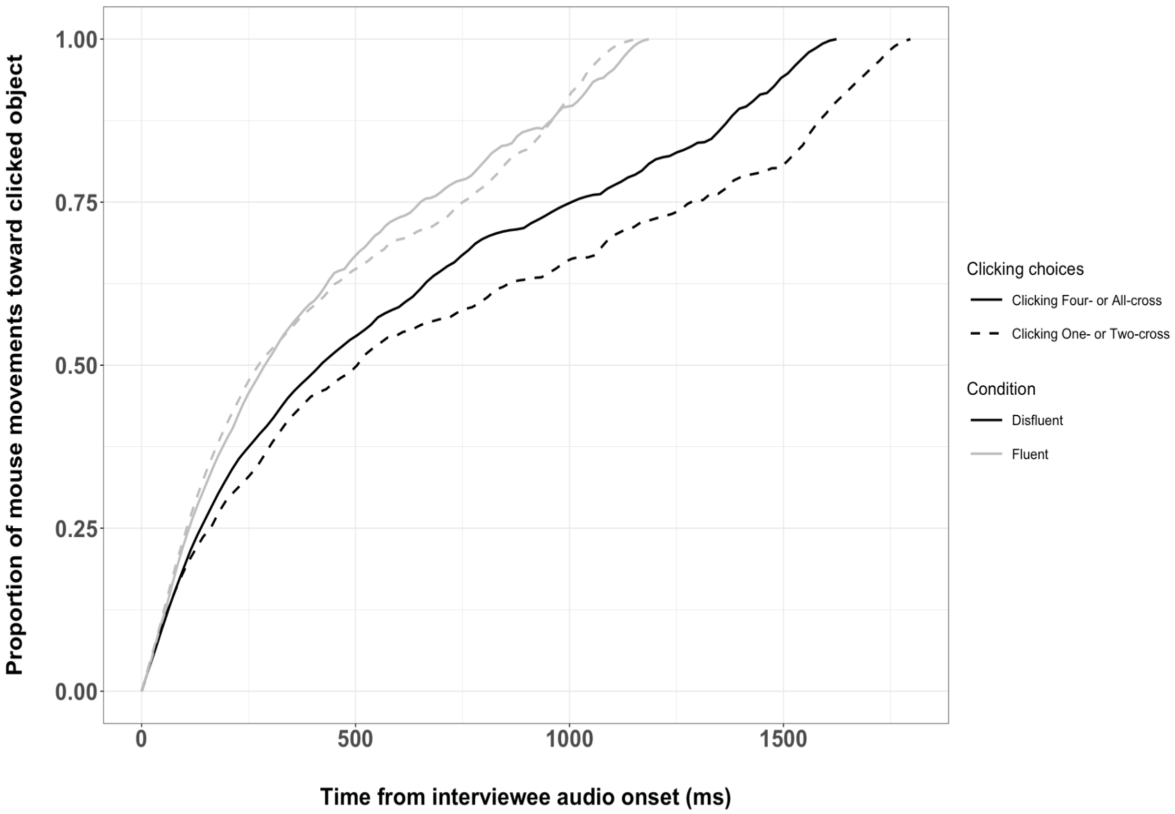


Figure 5.9 Proportion of mouse movements from the horizontal centre line of screen towards each group of clicking choices (larger-number group: four-/cross images; smaller-number group: one-/two-cross images) by condition (disfluent/fluent), starting from the onset of interviewee audio to the time when an image is clicked.

As shown in Figure 5.9, participants generally moved faster toward their choice of image in the fluent than disfluent condition. Following disfluent utterances, participants who clicked on images with larger numbers appeared to move faster compared to those clicking on smaller-number images. However, following fluent utterances, participants clicking different groups of images did not show obvious differences regarding their speed of movement.

This observed mouse-movement pattern was analysed statistically using BDOTS. The processed data were fit using a logistic function to compare the difference of curves for disfluent and fluent conditions (Top panel in Figure 5.10), regardless of where the final click images are. In the fitting stage, 38 curves had good fits without AR1 (R^2 >= .95), 2 curves had reasonable fits with AR1 (R^2 >= .80) and 116 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.992 and the adjusted alpha was calculated to be 0.00768. Regions of significance were found to start from time step 18 to time step 64 (Bottom panels in Figure 5.10).

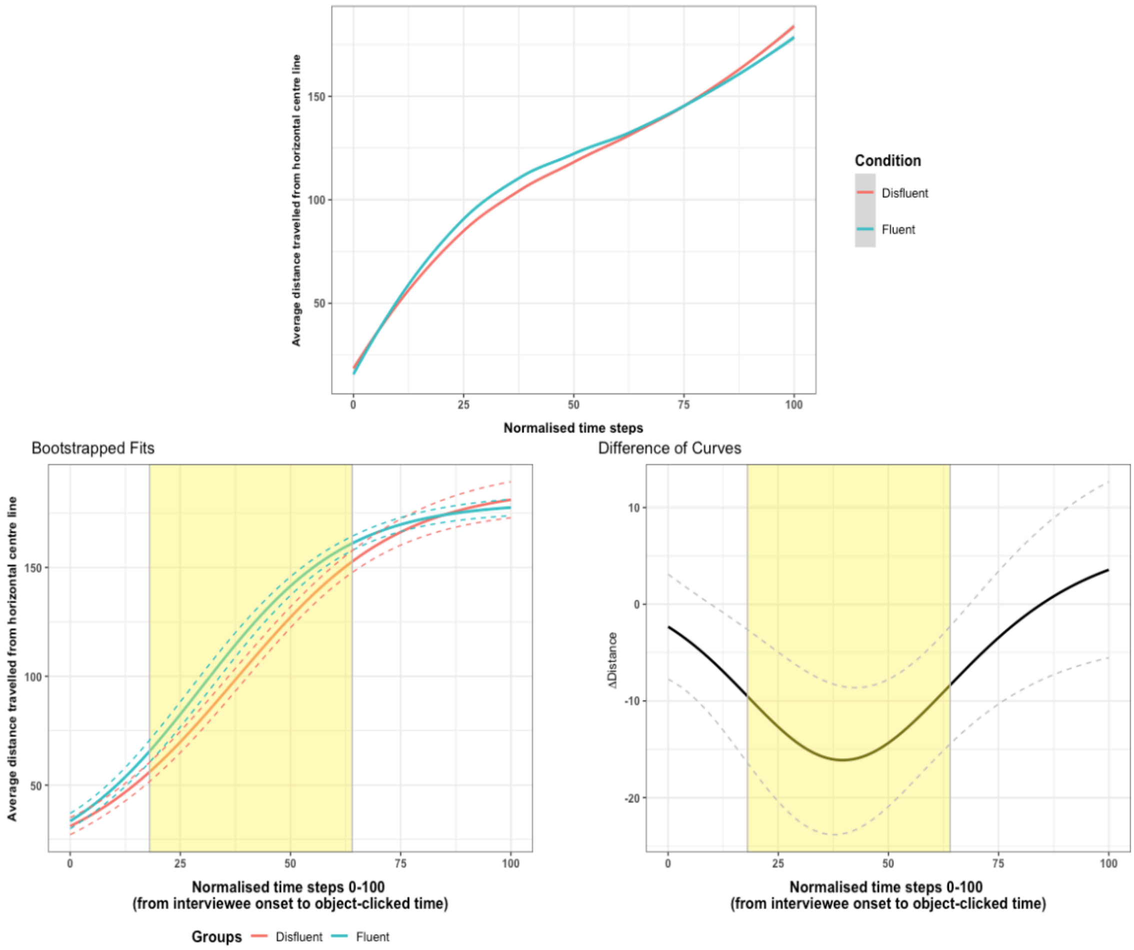


Figure 5.10 *(Top panel)* Two time-curves showing average distance (in pixels) from the mouse pointer to the horizontal centre line across 101 time-steps by manner of delivery (disfluent/fluent); *(Bottom panels)* BDOTS results of curve differences between conditions (disfluent/fluent) highlighting the significant time period of differences in Experiment 1.

This BDOTS analysis shows that, regardless of where the final click images are, participants were faster to move away from the centre line in the fluent compared to the disfluent condition. In particular, differences between conditions emerge at step 18 in the mouse trajectory. This result indicates that following disfluent utterances, listeners generally were less quick in deciding and moving towards a clicked image (less steep slope) compared to in the fluent condition. Critically, participants' mouse differences between disfluent and fluent conditions diverged early: 862ms after interviewee audio onset if mapping the time step 18 to actual timing of mouse movements.

Looking at cases with final clicks on images with larger numbers (Four-/All-cross images) in isolation, the mouse trajectories showed a similar effect of manner of delivery on participants' mouse movements. Data with final clicks on larger-number images were fit using a logistic function to compare the distance metrics between disfluent and fluent conditions (Top panel in Figure 5.11).

In the fitting stage, 6 curves had good fits with AR1 (R^2 >= .95), 36 curves had good fits without AR1, 3 curves had reasonable fits with AR1 (R^2 >= .80) and 90 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.994 and the adjusted alpha was calculated to be 0.00956. Regions of significance were found starting from time step 33 to time step 69 (Bottom panels in Figure 5.11).

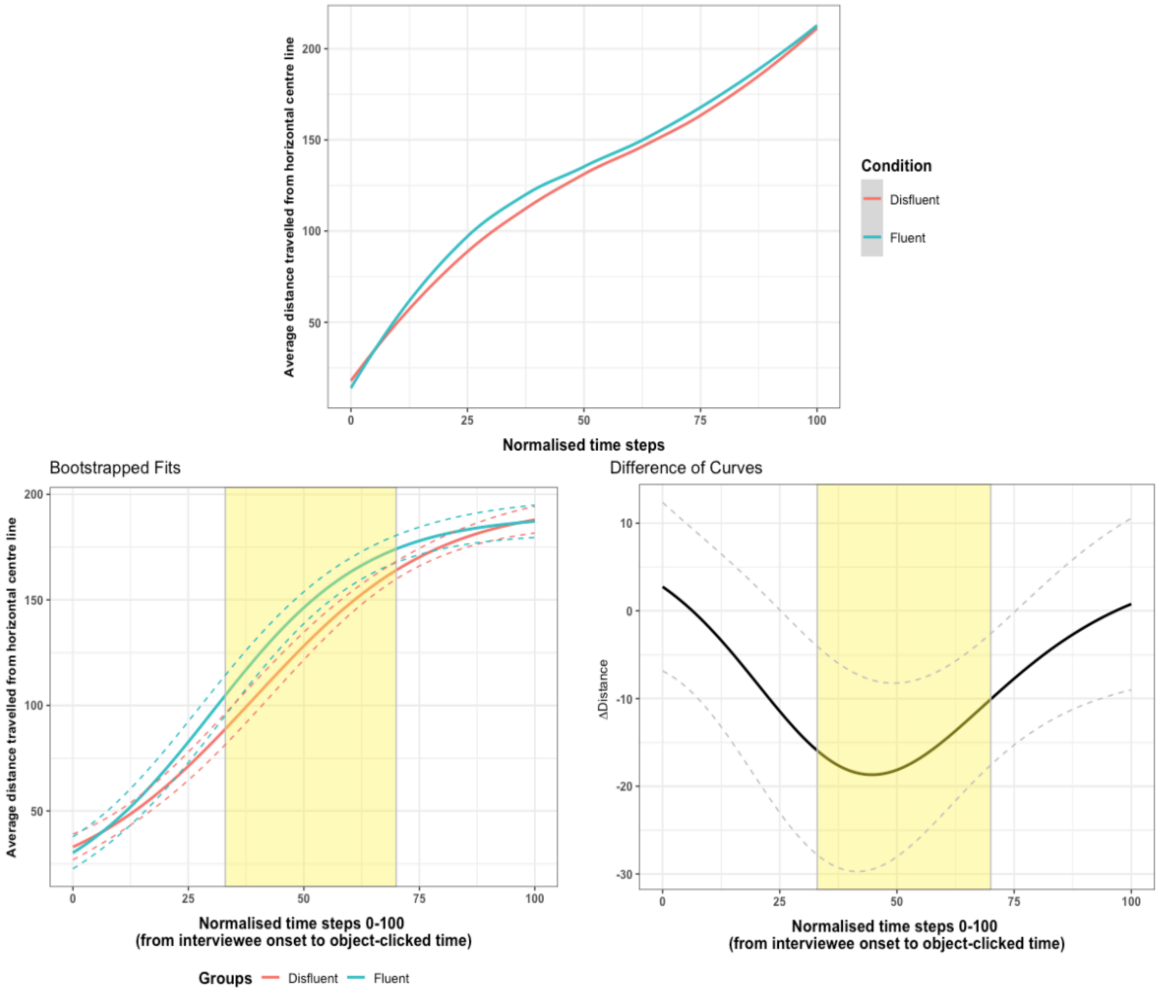


Figure 5.11 *(Top panel)* Two time-curves showing average distance (in pixels) from the mouse pointer to the horizontal centre line across 101 time-steps by manner of delivery (disfluent/fluent) in the cases where images with larger numbers (Four-/All-cross) were clicked; *(Bottom panels)* BDOTS results of curve differences between conditions (disfluent/fluent) when images with larger numbers (Four-/All-cross) were clicked, highlighting the significant time period of differences in Experiment 1.

Mapping this result into actual timing of mouse movements shows that for participants eventually clicking on images with larger numbers (Four-/All-cross images), their mouse movements between condition diverged quickly: around 1580ms after interviewee audio onset and critically 195ms after average disfluency *uh* onset in disfluent condition.

Exploring disfluent condition only, we further investigated the mouse trajectories dependent on where participants eventually clicked. Data in the disfluent condition were fit using a logistic function to compare the distance metrics between participants clicking on the images with larger numbers (Four-/All-cross images) and those clicking smaller-number images (One-/Two-cross images) (Top panel in Figure 5.12).

In the fitting stage, 18 curves had good fits with AR1 (R^2 >= .95), 59 curves had good fits without AR1, 10 curves had reasonable fits with AR1 (R^2 >= .80) and 187 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 1 and the adjusted alpha was calculated to be 0.03, showing that no significant regions were found where the two time-curves start to differ. This result (bottom panels in Figure 5.12) indicated that, following disfluent utterances, there was no significant different between participants' mouse movements toward images with larger numbers and toward images with smaller numbers.

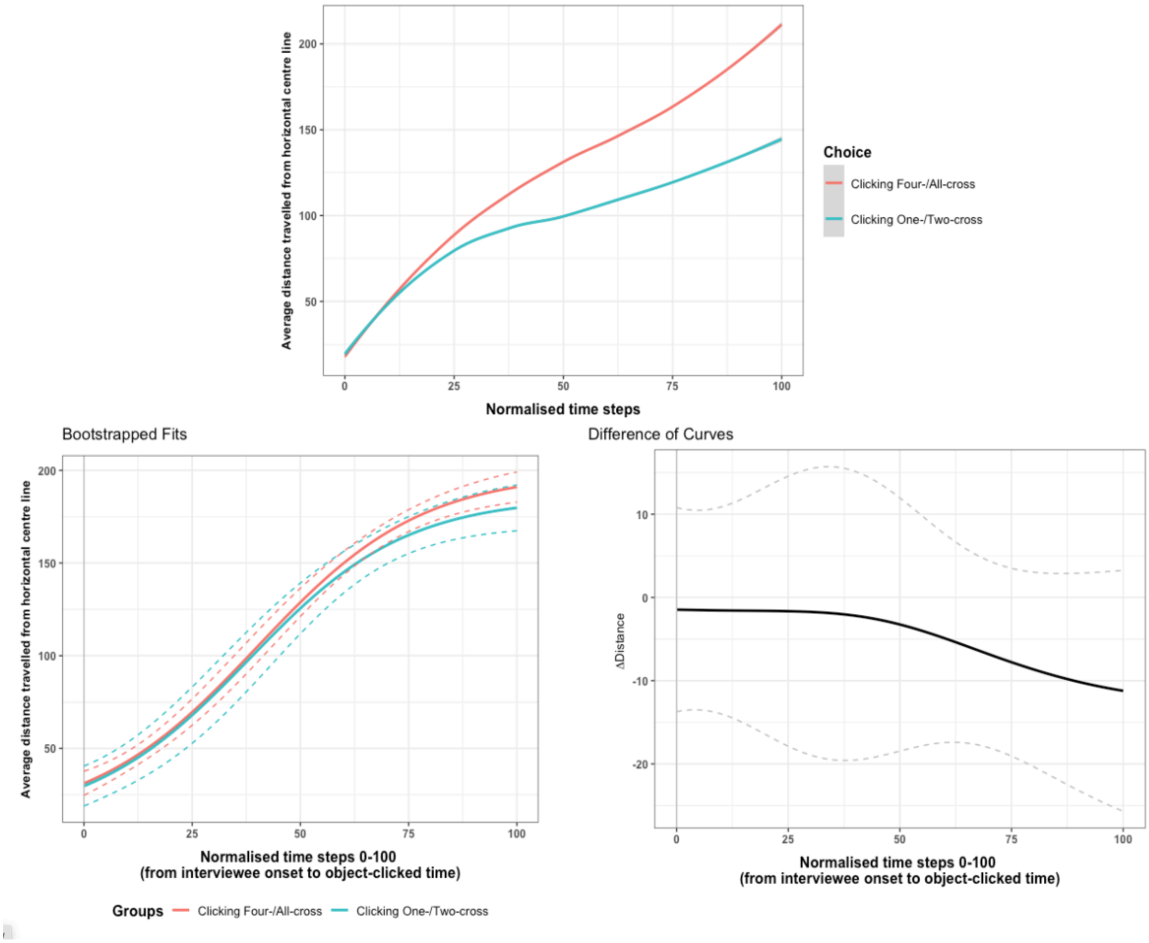


Figure 5.12 *(Top panel)* Two time-curves showing average distance (in pixels) from the mouse pointer to the horizontal centre line across 101 time-steps by whether the final click choice is larger-number group (Four-/All-cross images) or smaller-number group (One-/Two-cross images); *(Bottom panels)* BDOTS results of curve differences between final click choices on images with larger numbers (Four-/All-cross) and on images with smaller numbers (One-/Two-cross) in Experiment 1.

5.2.3 Discussion

Experiment 1 continues the research into how disfluency exerts its pragmatic effect on interpreting *some* based on previous studies, but we varied a social context and visual display to investigate whether the pragmatic role of disfluency retain. Specifically, we tested whether listeners still biased the undesirable interpretations of *some* following disfluent utterances in this context as shown in previous literature.

Experiment 1 provides further evidence for the pragmatic role of disfluency from two perspectives. First, in a different social context where more options of interpretations of *some* are presented, disfluency bias towards less socially favourable literal interpretations of *some* represented by images with a larger number of disqualifications. In 5.1% of disfluent and 2.4% of fluent trials, the meaning of *some* is stretched to the extreme upper end of a literal interpretation ("all"). Corresponding to results in Loy et al. (2019), the current results from Experiment 1 suggest that listeners take social context when processing scalar quantifier like *some* and disfluency is taken as cues of reasoning for the speaker's communication goals. Second, listeners integrate the pragmatic cues from disfluency into processing rapidly: their mouse movements diverged almost as soon as the audio onset between disfluent and fluent conditions.

However, even though "all" is a possible interpretation included in the upper bound of literal interpretation of *some*, extending *some* to this extreme end is an interesting finding. This result might be related to the nature of listeners' reasoning process cued by disfluency. Is this interpretation of *some* to "all" a speaker-modelling process of social reasoning under the given context, or is it a heuristic judgement of disfluency being a cue of lying and therefore any extreme values could be chosen if present? Experiment 2 aims to explore further this question.

5.3 Experiment 2

Experiment 1 shows that the speaker's manner of delivery has an influence on how listeners interpret the meaning of *some* in a given social context. However, the interesting finding of participants clicking on the extreme end of literal interpretations, all-crosses, makes us wonder the nature of listeners' reasoning process. One possibility is as shown in Loy et al. (2019), listeners applied a **speaker-modelling** action of social reasoning based on the given context: following disfluent utterance, listeners always bias the interpretations toward socially undesirable options, the upper end of literal interpretations in this case. The other possibility is listeners were actually merely heuristically taking disfluency as a cue of lying and any extreme values present to them could be chosen, no matter "all" or "zero". In this **heuristic** account, the meaning of *some* is expected to be ignored by listeners and disfluencies act as the main determiner of the interpretations of utterances.

To distinguish between a speaker-modelling account or a heuristic account of comprehension in this context, Experiment 2 replicated Experiment 1 but changing only one aspect of the design. In Experiment 2, the lower-bound meaning “one” was replaced by a “zero”. As a meaning that is not compatible with the meaning of *some* at all, this design was able to explore whether the role of cue-of-lying was present in listeners’ processing in this social context. If listeners were reasoning following a speaker-modelling account, it is still expected to see more interpretations of *some* on the upper end of literal interpretations as in Experiment 1. However, if disfluency is merely a heuristic cue of lying, participants may be observed to have clicks on zero-cross image, which also represents an extreme end of numerical value in this context.

5.3.1 Methods

5.3.1.1 Participants

We recruited participants with the same standard as for Experiment 1. Mouse movement data from a hundred and sixty eight (Female: 80 ; Male: 87 ; Non-binary: 1) self-reported native English speakers recruited via Prolific Academic were analysed further, with data from an additional 12 participants ruled out (filtered with the same standard in Experiment 1). Participants provided informed consent by clicking the ‘I consent’ button on information page following the University of Edinburgh Research Ethics Committee guidelines (reference number: 271-2122/4).

5.3.1.2 Material and experiment design

As in Experiment 1, we recorded participants’ mouse movements in a web-based task in which we manipulated Disfluency (present vs. absent) within-subjects in a set of target trials (N=12). For each target trials, participants saw four images with four different counts of qualifications displayed on the screen (see Figure 5.13 for example trials) and heard recorded interview conversations between an interviewer and one of six interviewees. Experiment 2 was implemented in jsPsych version 7.0 (De Leeuw, 2015).

Figure 5.13 Display examples for four target trials in Experiment 2 (Top-left panel: interviewer asking "How many times have you been late for a scheduled delivery?" and interviewee answering "I have been late for *[uh] some* of the scheduled deliveries"; Top-right panel: interviewer asking "How many times have you had complaints from customers recently?" and interviewee answering "I've had *[uh] some* complaints"; Bottom-left panel: interviewer asking "How many 'F's did you get for driving proficiency tests?" and interviewee answering "I got *[uh] some* 'F's"; Bottom-right panel: interviewer asking "How many inspections has your vehicle failed over the past few months?" and interviewee answering "My vehicle failed *[uh] some* inspections.")

5.3.1.3 Audio and visual stimuli

Experiment 2 used the audio stimuli from Experiment 1. For the visual stimuli, participants saw four images from one set with zero, two, four or all (five) disqualifications marked in red. As in Experiment 1, the allocation of four images was fixed (with zero-cross on the top left, two-cross top right, four-cross bottom left and all-cross bottom right).

5.3.1.4 Procedure

Other than the use of zero-cross in place of one-cross images, participants experienced an identical procedure of Experiment 1 in Experiment 2.

5.3.2 Results

The data from Experiment 2 was analysed the same way as in Experiment 1 in R version 4.2.3 (R core team, 2023). Filler trials, practice trials, and attention-check trials were removed prior to analysis.

5.3.2.1 Final object click

Click results

Figure 5.14 shows the total number and the percentages of final clicks on each group of images (larger-number group: four-/all-cross images; smaller-number group: zero-/two-cross images) by manner of speech (disfluent/fluent).

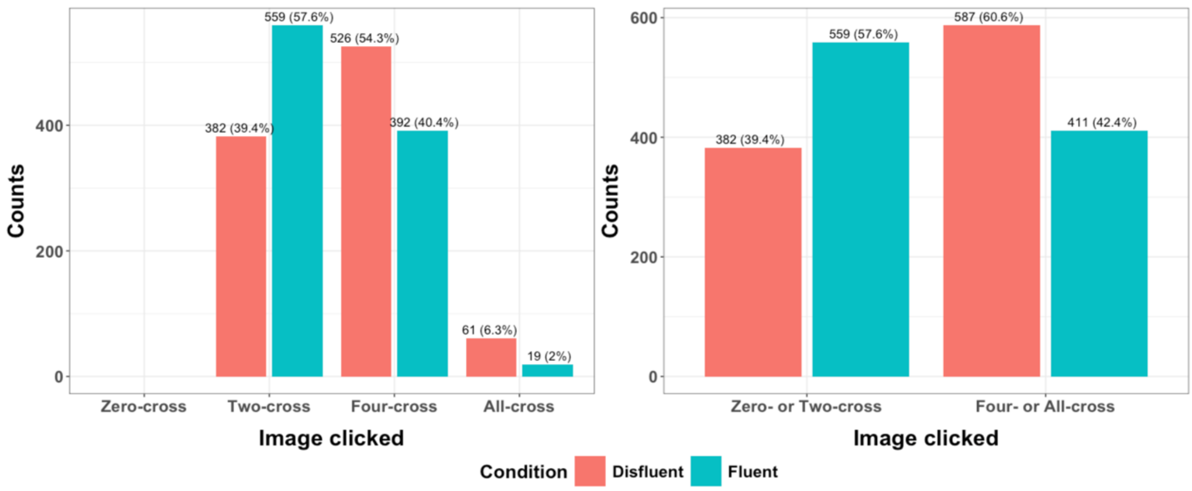


Figure 5.14 *(Left panel)* Bar plot showing total number and distribution of mouse clicks recorded on each image (Zero-cross, Two-cross, Four-cross, or All-cross) by manner of delivery (disfluent/fluent); *(Right panel)* Bar plot showing total number and distribution of mouse clicks recorded on each group of choices (images with smaller numbers: Zero-/Two-cross; with larger numbers: Four-/All-cross) by manner of delivery (disfluent/fluent).

Similar to clicking results in Experiment 1, participants' responses showed a general bias towards larger-number interpretation cued by disfluency. Following disfluent compared to fluent utterances, participants clicked more on larger-number choices (60.6% vs. 39.4%) and less on smaller-number choices (42.4% vs. 57.6%). Critically, there were no clicks on the zero-cross image in either the disfluent or fluent condition. As in Experiment 1, where clicks on the larger-number images increased by 17.8% following disfluency, participants' clicking choices are influenced by the speaker's manner of delivery ($\beta = 1.205$, $SE = 0.178$, $p < 0.001$). The fact that there was no click on the zero-cross image indicated the heuristic account that listeners always heuristically relate disfluency to lying fails to hold. Participants' response pattern corresponds to the speaker-modelling account that listeners reason contextually for the speaker's disfluency.

Time taken to click

The overall time participants took to click their final pick of image (larger- and smaller-number images) after the onset of *some* is shown in Figure 5.15. As in Experiment 1, participants' time taken to click is not influenced by whether the speaker's disfluency ($\beta = 57.5$, $SE = 78.2$, $t = 0.74$) or by an interaction between the speaker's disfluency and where they clicked eventually ($\beta = -196.9$, $SE = 113.7$, $t = -1.73$), but dependent on whether they clicked on images with larger numbers. Participants eventually clicking on smaller-number (Zero-/Two-cross) images took significantly longer time compared to those clicking on larger-number (Four-/All-cross) images ($\beta = 351.1$, $SE = 89.6$, $t = 3.92$).

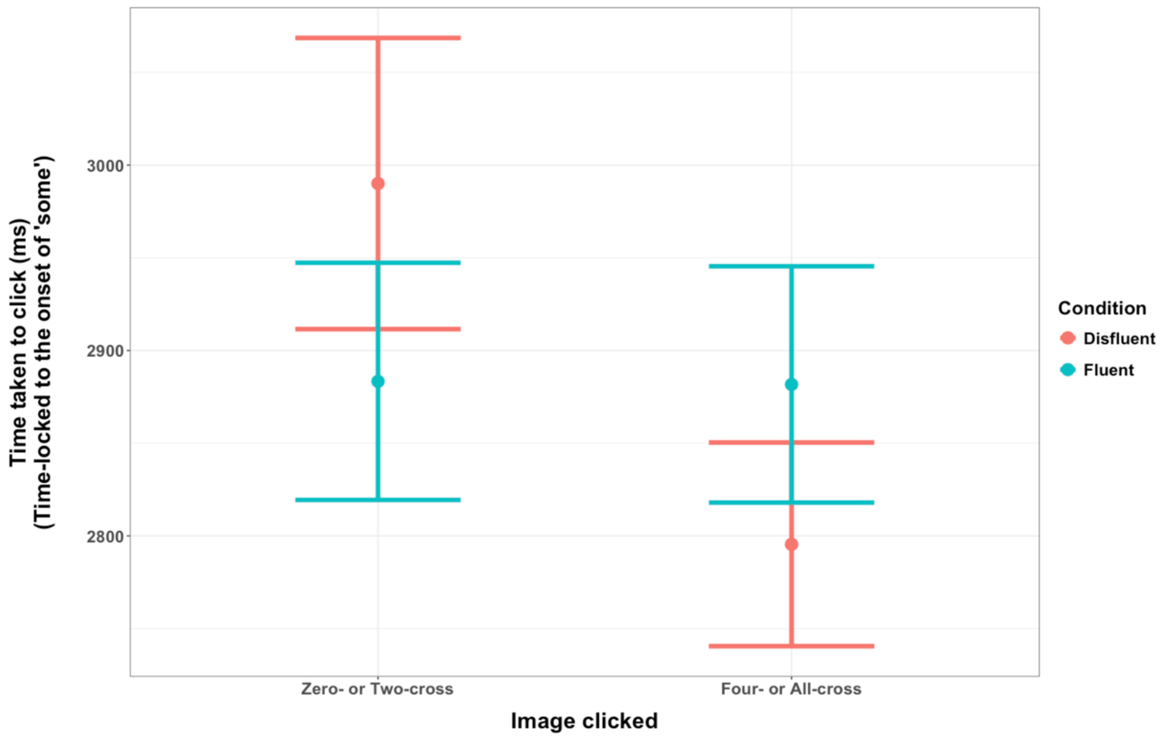


Figure 5.15 Average time used in clicking (relative to the onset of critical word *some*) on each group of choices (images with smaller numbers: Zero-/Two-cross; with larger numbers: Four-/All-cross) by manner of delivery (disfluent/fluent). Error bars represent ±1 SE of the mean clicking time.

5.3.2.2 Mouse movements

We processed the collected mouse-tracking data from all 167 participants following the same steps as in Experiment 1. As shown in the clicking results, replacing One-cross image to Zero-cross image appeared not to influence participants' mouse movements at all. The proportion of participants' mouse movements toward their clicked object across time in disfluent and fluent condition is shown in Figure 5.16. People generally moved faster toward a clicked image in fluent condition. However, in disfluent condition only, people moved faster towards larger-number images (Four-/All-cross) when they were going to compared to those clicking smaller-number images (Zero-/Two-cross). This pattern of participants mouse movements replicated that in Experiment 1, indicating that disfluency slowed people down in deciding an image to click, but only following disfluent utterances, did participants show a faster movement towards larger-number interpretation of *some*.

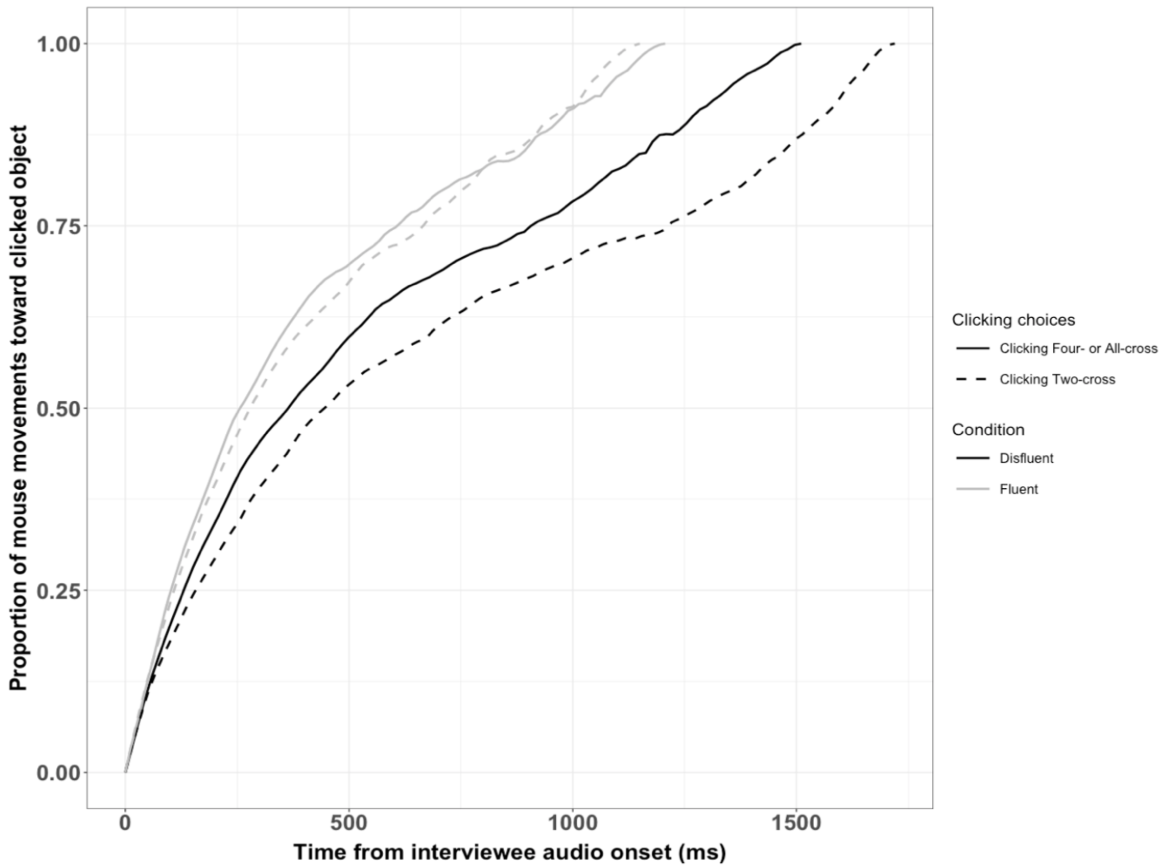


Figure 5.16 Proportion of mouse movements from the horizontal centre line of screen towards each group of clicking choices (larger-number group: four-/cross images; smaller-number group: zero-/two-cross images) by condition (disfluent/fluent), starting from the onset of interviewee audio to the time when an image is clicked.

5.3.2.3 Mouse trajectories

Following the analysis steps in Experiment 1 again, I calculated the distance between mouse pointers to the horizontal centre line to indicate participants' movements away from a neutral line. Obtained data were fit using a logistic function in a BDOTS analysis. I compared the difference of time-curves for disfluent and fluent conditions across all items first (Top panel in Figure 5.17).

In the fitting stage, 10 curves had good fits with AR1 (R^2 >= .95), 59 curves had good fits without AR1, 5 curves had reasonable fits with AR1 (R^2 >= .80) and 116 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.998 and the adjusted alpha was calculated to be 0.0162. The result that no significant region was found where these two time-curves start to differ (Bottom panel in Figure 5.17) suggests that in Experiment 2 where zero-cross replaces one-cross image, the timing of participants' choice towards which group of images to choose was not significantly influenced by the speaker's manner of delivery.

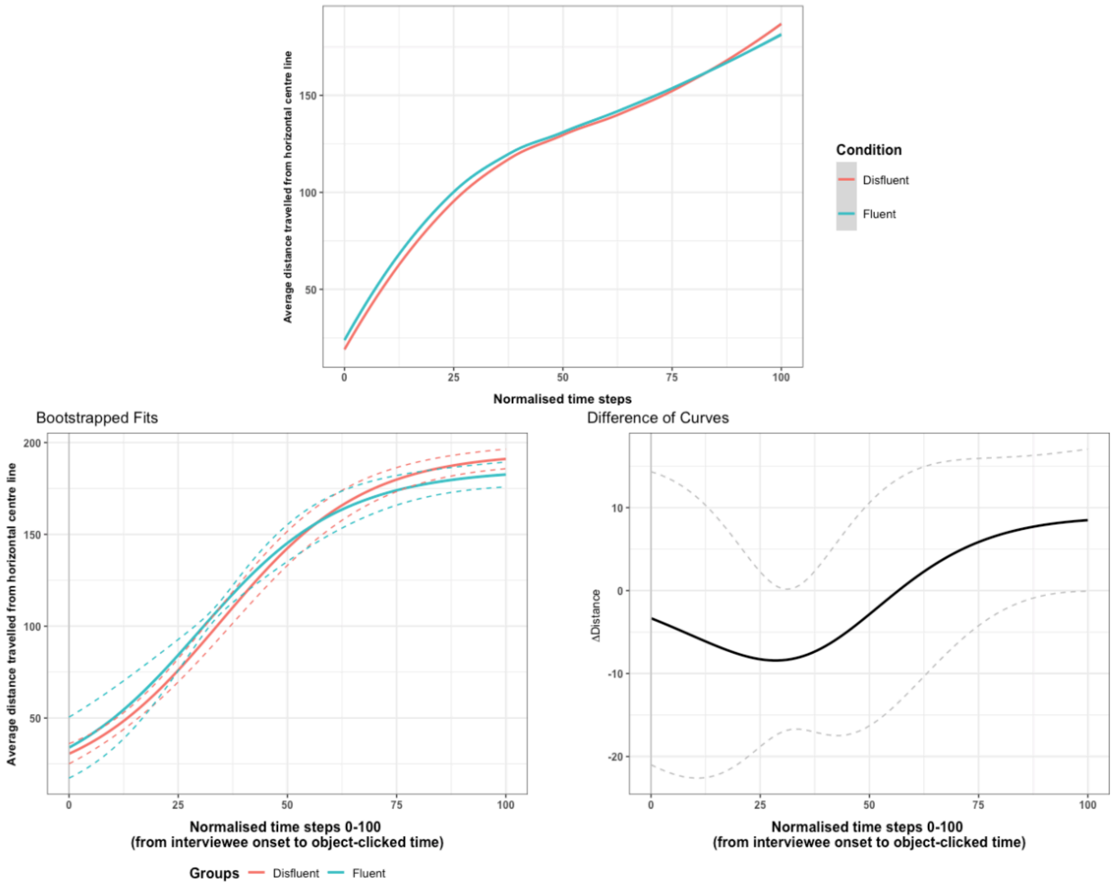


Figure 5.17 *(Top panel)* Two time-curves showing average distance (in pixels) from the mouse pointer to the horizontal centre line across 101 time-steps by manner of delivery (disfluent/fluent); *(Bottom panels)* BDOTS results of curve differences between conditions (disfluent/fluent) in Experiment 2.

Focusing on the disfluent condition only, a second analysis was performed. The data was fit using a logistic function, to compare the curve differences between participants clicking on the larger-number images (Four-/All-cross) and those clicking on smaller-number images (Top panel in Figure 5.18).

In the fitting stage, 28 curves had good fits with AR1 (R^2 >= .95), 94 curves had good fits without AR1, 7 curves had reasonable fits with AR1 (R^2 >= .80) and 172 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 1 and the adjusted alpha was calculated to be 0.0325. Again, there showed no significant time period of differences (Bottom panel in Figure 5.18), indicating that following disfluent utterances, participants' mouse-movements towards either group of images do not differ significantly over time.

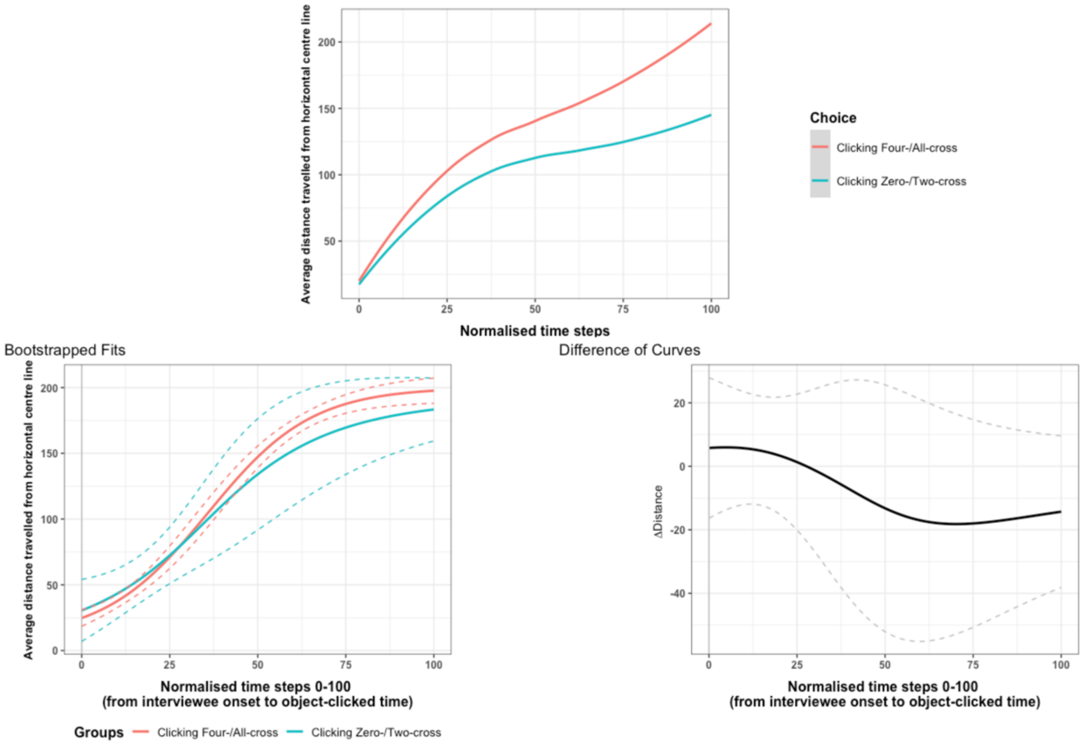


Figure 5.18 *(Top panel)* Two time-curves showing average distance (in pixels) from the mouse pointer to the horizontal centre line across 101 time-steps by whether the final click choice is larger-number group (Four-/All-cross images) or smaller-number group (Zero-/Two-cross images); *(Bottom panels)* BDOTS results of curve differences between final click choices on images with larger numbers (Four-/All-cross) and on images with smaller numbers (Zero-/Two-cross) in Experiment 2.

5.4 General discussion

The current study investigates how a speaker's manner of delivery influences listeners' real-time processing of the ambiguous scalar quantifier *some*. By varying the context and paradigm used in Loy et al. (2019), the current two web-based mouse-tracking tasks were to test whether the pragmatic role of disfluency in processing *some* retains, and to explore further the nature of this contextual effect of disfluency: whether it is a speaker-modelling action of social reasoning or a heuristic connection of disfluency and lying. The mouse-tracking results reported here show a similar effect of disfluency on listeners' interpretations of *some*, relating disfluency to a socially undesirable meaning: participants are more likely to select images with larger values of disqualifications following disfluent utterances. These response patterns are compatible with the suggestion that disfluency acts as a pragmatic cue in a social context with *some* made by Loy et al. (2019).

Moreover, this pragmatic effect of disfluency is achieved via listeners considering speaker's communication goals rather than listeners heuristically relating disfluency to a speaker's lying behaviour. Taken together the clicking patterns in both experiments, not choosing the lower-end extreme numeric value "zero" at all (in Experiment 2) suggests that participants were not choosing the extreme upper bound meaning of *some*, "all" merely via heuristically relating disfluency to lying, but reasoning according to the given context.

Regarding the time-course of processing, differences in participants' mouse movements between disfluent and fluent conditions show that listeners take the contextual information involved in disfluency rapidly, almost as soon as the onset of audio. For those eventually clicking on larger values, their mouse movements following disfluent utterances significantly diverge from the movements in fluent utterances rapidly after the disfluency onset.

The significant time differences between people clicking on images with larger numbers and those clicking on smaller-number images indicates that in this given context, listeners are faster to move to a larger-value compared to a smaller-value. This pattern corresponds to the choice bias discussed in previous literature (Bonnefon et al., 2009; Feeney & Bonnefon, 2013), indicating again that in a social context with a potential facial loss involved, people are biased toward a more socially undesirable interpretations of ambiguous scalar words.

This study adds up to the findings of Loy et al. (2019) by providing a scale of interpretations of *some*, as well as applying a direct display to avoid noise in cognitive effort on processes other than interpreting *some*. More importantly, a conceptual replication of a previous visual world paradigm in Loy et al. (2019), the current study shows the possibility of applying a mouse-tracking-only paradigm in exploring real-time language processing via an easy implementation and quick recruitment of participants. Also, with four images presented in four quadrants of the screen display, the current study applies a more complicated mouse-tracking paradigm compared to the commonly used two-image mouse-tracking paradigm. In this way, more trajectory analysis methods are encouraged to be explored for studying the time-course of processing.

The two mouse-tracking experiments suggest a pragmatic effect of disfluency in facilitating listeners' processing of ambiguous scalar quantifier via modelling the speaker's communication goals in a given context. The current findings leave room for exploring the effect of disfluency in different social contexts and in processing of other ambiguous scalar words, such as *a few*, *likely*, etc.

5.5 Conclusions

This chapter reported two mouse-tracking experiments investigating the effect of disfluency in listeners' real-time processing of ambiguous scalar quantifier *some*. Following disfluent utterances, listeners associate some with larger values, showing a bias towards a socially undesirable meaning of *some* in the given context. Even though the current result corresponds to the findings in Loy et al. (2019), questions remain about whether this bias towards larger-value literal interpretation is indeed a pragmatic effect of disfluency, or it is merely an association between disfluency and a literal interpretation of *some* in any context. Experiment 3 and 4 (Chapter 6) investigates further on the nature of this social reasoning process cued by disfluency's pragmatic effect.

Chapter 6

When *some* is less: Disfluency in context

6.1 Introduction

When interpreting an utterance in daily conversation, listeners not only take into consideration what is said, but also *how* it is said. For example, disfluencies produced in spontaneous speech such as fillers like *uh* or *um* have been shown to affect the inferred meaning of an utterance (Bonnefon, Dahl, & Holtgraves, 2015; Loy, Rohde, & Corley, 2017). However, it is not yet clear how listeners alter their unfolding interpretation of an utterance when encountering disfluency. In particular, it is not clear whether disfluency has an attentional effect or whether context interacts with disfluency in processing to affect interpretation more directly. Following the findings in Chapter 5, we introduce two additional web-based mouse-tracking experiments here. These two experiments are designed to further evaluate the influence of context on listeners' online interpretations of ambiguous statements over time. In this chapter, our first experiment establishes that context has a specific role to play in the interpretation of disfluent utterances, and our second investigates the way in which meaning is established when these utterances are disfluent: Whether listeners take disfluency as a sign of speaker deception, or as a sign of a more subtle speaker tactic of communicating a true but strategically selected meaning of an utterance.

The experiments capitalise on the broad ambiguity represented by the scalar quantifier *some*. Consider the following exchange (1) between two speakers:

(1) A: "How many oreos have you eaten?"
B: "I have eaten some oreos."

Potential interpretations of B’s answer vary from ‘I have eaten some and possibly all of the oreos’, the semantic meaning, to ‘I have eaten some but not all of the oreos’, the pragmatically-strengthened meaning. Importantly, these interpretations of *some* depend on linguistic context (Breheny, Katsos, & Williams, 2006; Politzer-Ahles & Husband, 2018). People’s preference for the pragmatically-strengthened interpretation varies depending on the preceding context: Self-paced-reading results show that the ‘some but not all’ interpretation is favoured when it is in answer to a question such as “Have all the students passed the exam?” compared to when it is used to answer “How many students have passed the exam?” (see Politzer-Ahles & Husband, 2018). With the former question, the focus on whether ‘all’ is true facilitates the readers’ understanding of the pragmatic, ‘not-all’, interpretation of the meaning of *some*.

Besides the linguistic context, the interpretation of *some* may also depend on social context and the manner of delivery. In a study using written vignettes describing spoken interactions, Bonnefon et al. (2015) suggested that comprehenders could take silent pauses as a cue which shifts expectations towards a less socially desirable meaning. The study manipulated whether or not a silent pause was included before a speaker uttered a face-threatening expression (see an example in (2) below) and asked participants to judge the speaker’s meaning, by rating the likelihood that the reply endorsed the socially undesirable meaning (i.e., all people hated your idea).

(2) Yesterday, you pitched an idea to a group of five persons. Today, you ask Bob (who was in the group) what people thought of your idea. Bob <*stays silent for a few seconds*>. Then he replies: “Some people hated your idea.”

The socially undesirable meaning was rated as more likely when the speaker was described as remaining silent before speaking, perhaps because participants inferred that a speaker may wish to convey to the listener the more charitable ‘some but not all’ meaning, but their silence suggests that this strategic use of *some* is concealing a less desirable ‘some and in fact all’ meaning. Bonnefon et al.’s finding indicates that within a relevant context, the eventual interpretation of *some* depends on the manner in which the speech is delivered.

The effects that Bonnefon et al. (2015) found hold for online processing of spoken language as well. Manipulating the speaker’s manner of speech, Loy et al.

(2019) used an eye-tracking paradigm to establish that, in a situation where ascribing higher values to *some* is socially undesirable (‘I ate some and possibly all oreos’), listeners were more likely to make an early commitment to the semantically-licensed interpretation that all the oreos had been eaten when the speaker was disfluent (“I ate, uh, some oreos”).

Although Bonnefon et al. and Loy et al. accounted for their findings in terms of social desirability, the tendency to interpret *some* as more likely to mean ‘all’ could have arisen from a simple attentional bias. Disfluencies have been found to heighten listeners’ attention, affecting how listeners process spoken language. Listeners pay special attention to and are more likely to recognise having heard the words preceded by disfluencies (Corley, MacGregor, & Donaldson, 2007; Collard, Corley, & MacGregor, 2008). Following disfluency, listeners’ attention may be oriented to the core, literal meaning of the word *some*, which would make the semantic interpretation of *some* more salient. If this *heightened attention* account holds, the effect of disfluency on listeners’ interpretations of the word *some* should always favour the semantic interpretation, regardless of the given context.

However, if Bonnefon et al. and Loy et al. are correct in that spoken manner influences listeners’ pragmatic interpretation of utterances, we might expect to see people understand *some* differently depending on the *social reasoning* associated with different contexts. Note that Bonnefon et al.’s (2015) and Loy et al.’s (2019) studies used social contexts where the semantic meaning (‘I have eaten some and possibly all oreos’) is also the socially undesirable meaning, and these studies therefore cannot distinguish the attentional and social accounts. The present study introduces a different social context where, critically, the socially undesirable meaning is now the pragmatically-strengthened interpretation of *some* and the semantic ‘some and possibly all’ meaning is thus expected to be the one that a speaker would prefer to convey. The goal is to disambiguate the two posited accounts of listeners’ interpretation of *some* (a simple bias arising from heightened attention versus context-sensitive social reasoning).

6.1.1 The present study

We measured the online interpretation of *some* using a mouse-tracking technique in a web-based task. The social context used in the current two experiments was a job interview for a research assistant position. Specifically, we used as a set-up a series of recorded interviews where participants heard conversations like (3), which manipulated whether the interviewees were disfluent or not when answering ("I've got some 'A's" vs. "I've got, uh, some 'A's").

(3) Interviewer: "How many 'A's have you got for your psychology modules?"
Interviewee: "I've got *some* 'A's." (fluent) *or*
"I've got, *uh, some* 'A's." (disfluent)

In each of two web-based clicking tasks, participants saw four images on the screen with different numbers of interviewee qualifications (represented by numbers of 'A' grades, or number of ticks for other qualifications) while hearing the interview audios. The qualifications refer to things that are desirable in the given context of job interview, such as the number of 'A' grades or the number of languages one can speak. In Experiment 3, one of the images depicted one qualification as an available option for interpretation. Even though 'one' is not a usual interpretation of *some*, it maps on to the lower-bound meaning of *some* (i.e., 'at least one'). This number was included to test whether it was possible for participants to extend the meaning of *some* to its absolute lower bound ('some but not all, in fact only one'), given a context that supports the relevant social reasoning. For example, in the current context where more qualifications are favoured, an interviewee might use *some* as a potential 'vague exaggeration' of the one 'A' qualification they have actually attained.

We recorded participants' mouse movements throughout the experiment session, measuring for each trial both their clicks (i.e., which image was clicked) and these trajectories of the mouse movements made before each click. If the *heightened attention account* predicts the real-time processing of *some*, we expect to see results similar to those of Loy et al. (2019) and to those in Chapter 5, with preference towards images compatible with the semantic 'some and possibly all' interpretation following disfluency, despite the use of an interview context with a different social bias. However, if listeners' interpretations of *some* are driven by *social reasoning*, participants are expected to interpret disfluency as indexing the less desirable interpretation of *some* according to the specific social context. In our case, the pragmatic 'some but not all' interpretation is less desirable in the interview context. Accordingly, under this account, we would predict more final clicks on and early mouse moving towards images depicting fewer qualifications following disfluent compared to fluent utterances.

6.2 Experiment 3

6.2.1 Method

6.2.1.1 Participants

One hundred and fifty (Female: 90, Male: 57, Non-binary: 3) self-reported native English-speakers were recruited via Prolific Academic, and each participant was compensated £2 for a task that was estimated to take 15 minutes (£8/h). An additional 29 participants were tested but their data were not included in our analyses, because these participants (1) failed to finish the whole experiment session, (2) took more than 40 minutes to finish the whole session (this time limit is based on pre-tests with non-native English speaker participants), or (3) failed one or both attention checks. All participants were between 18 and 35 years old, with normal or corrected-to-normal vision, and with no self-reported hearing difficulties. Every participant provided informed consent by clicking a button on an information page at the beginning of experiment, in accord with the University Research Ethics Committee approval (reference number: 84/2021-6).

6.2.1.2 Material and experiment design

We conducted a within-subjects experiment, manipulating the presence of disfluency (present vs. absent) in critical trials. Each participant listened to 6 interviews, each comprising 6 exchanges between an interviewer and one of 6 interviewees. Each exchange constituted one trial, and was associated with a set of four images (Figure 6.1). All of the interviews comprised exchanges on the same six topics, using the same 6 sets of four images. Within each interview, 2 of the exchanges were critical trials; one each with a fluent or disfluent response from the interviewee. Over the six interviews, each type of qualification served once as a critical disfluent and once as a critical fluent item, so that each participant encountered 12 critical trials (6 disfluent).

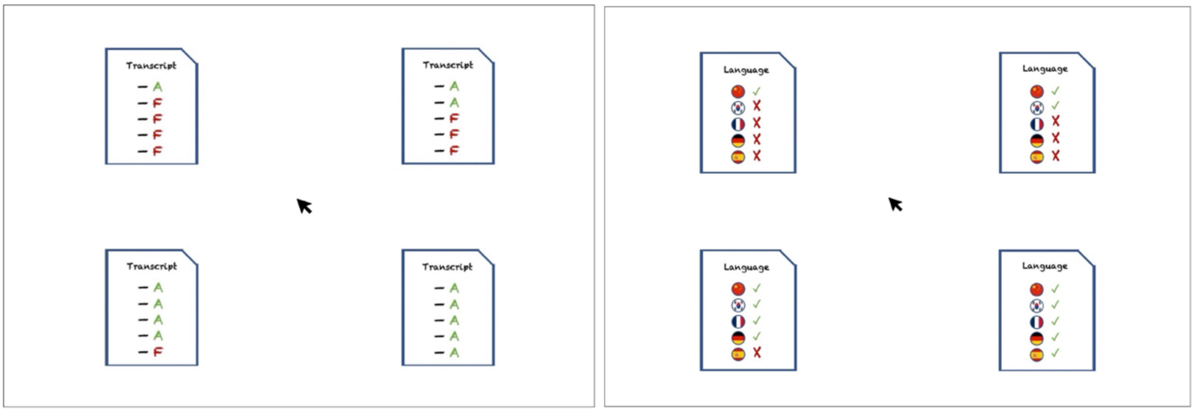


Figure 6.1 Displays for two target trials in Experiment 3 (Left panel: interviewer asking “How many ‘A’s have you got for your psychology modules?” and interviewee answering “I’ve got *[uh] some* ‘A’s”; Right panel: interviewer asking “How familiar are you with the listed languages?” and interviewee answering “I can speak *[uh] some* of them.”)

6.2.1.3 Visual stimuli

Six sets of images indicating counts of qualifications corresponding to six interview questions were used as potential referents in the experiment. The six questions concerned the following qualifications: Q1: Number of good grades in the psychology modules in the university; Q2: Number of the training courses taken before application; Q3: Number of criteria the participant meets for the standard of the application; Q4: Number of languages the participant is familiar with; Q5: Number of the lab’s previous projects the person is familiar with; Q6: Number of completed recommended reference books.

In each target trial, participants saw four images with one, two, four, or all (five) qualifications – number ‘A’s versus ‘F’s (green versus red) or number of ticks marked in green with the remainder marked with a cross in red (as the right panel in Figure 6.1). In one case, green “A”/red “F” was used (as in Figure 6.1). For convenience, these images will be further referred to as one-tick, two-tick, four-tick and all-tick. In response to each interview question, more ticks or ‘A’s in the images represent a higher number of qualifications. Images always appeared in fixed locations (one-tick top left, two-tick top right, four-tick bottom left, all-tick bottom right).

6.2.1.4 Audio Stimuli

For every participant, the whole experiment session included six complete interviews. Every complete interview comprised six conversations between the interviewer and an interviewee, where two of the six conversations were critical trials and the remaining four were filler trials, used to reduce the chance that participants would notice the experimental design. In total, every participant heard 36 exchanges, including 12 critical and 24 filler ones.

Recording of audio stimuli was approved by the University Ethics Committee with the reference number 119/2021-2. Six native British English speakers (three males and three females) recorded the interviewee scripts, and one native American English speaker recorded the interviewer script.

In critical trials, the recorded interviewees' answers included *some* in either the disfluent or fluent condition. Interviewees' answers in filler trials contained no use of the disfluency *uh* or the word *some*, but included other forms of disfluency (e.g., repetition, prolongation, 'like', 'hmph', etc.) and other expressions of amount[15] (e.g., 'almost', 'most of', 'more/less than', 'at least', 'around', 'a few', 'approximately', 'not much', etc.). For example, the interviewee's answer in a filler trial about the number of training courses taken is "Actually, due to the time limit, I, I, I only took a few of them", using repetition as a disfluency indicator and 'a few' to indicate the amount.

The interviewer and interviewee recordings were made online via a browser-based recording tool implemented in JavaScript.[16] Step-by-step instructions were provided to speakers on how to use the recording website, with details to pay attention to before and during recording. Speakers were asked to record each sentence at least three times, which made it possible for the experimenters to choose the most natural utterances.

The collected recordings were further edited using Audacity (version 2.4.2). In two cases, fluent and disfluent recordings sounded quite different (with obvious intonation differences), and a disfluent *uh* was inserted into a fluent recording; in two cases the *uh* was excised from disfluent recordings to produce the fluent recordings participants had forgotten to record.

15 As in Experiment 1-2 in Chapter 5, fillers did not mention specific numbers.
16 Recordings were made during the Covid-19 pandemic under UK social distancing regulations.

The remaining recordings were all edited in the following way: for fluent utterances, the most natural version was chosen and used; for disfluent utterances, the most natural disfluency *uh* was extracted from one disfluent utterance and was added to the other disfluent utterance (replacing the original *uh*), so that the two critical trials in the same interview had the same disfluency *uh*. For fillers, no editing was applied.

6.2.1.5 Procedure

Participants completed the experiment, implemented in JavaScript using jsPsych (De Leeuw, 2015) online. Once they had read the welcome page, they clicked the start button and answered five demographic questions. The main mouse-tracking experiment session started with a trial in which the audio instructed participants to click on a specific image. This was designed to familiarise participants with the experiment, as well as to test that they could easily understand the audio, leaving time for them to adjust the volume as required. Following the audio check, a brief instruction page reviewed the cover story for the study and explained the experimental procedure.

The flow of the experiment is shown in Figure 6.2. At the start of each trial, participants clicked a central Ready button, which ensured that the pointer was at the centre of the screen. After clicking the Ready button, four images appeared on the screen, corresponding to four possible interpretations of the interviewee's answer. After 2000ms, the first part of the audio (interviewer question) was played. During this part of the audio, participants could move the pointer using the mouse, but clicks (on any of the four images or elsewhere) were not registered.

Once the interviewee's response started playing, participants were able to click on any of the four images as soon as they wished. If no click was made within 10s of the interviewee audio offset, the trial timed out and the screen automatically showed the 'Ready' button indicating the start of the next trial.

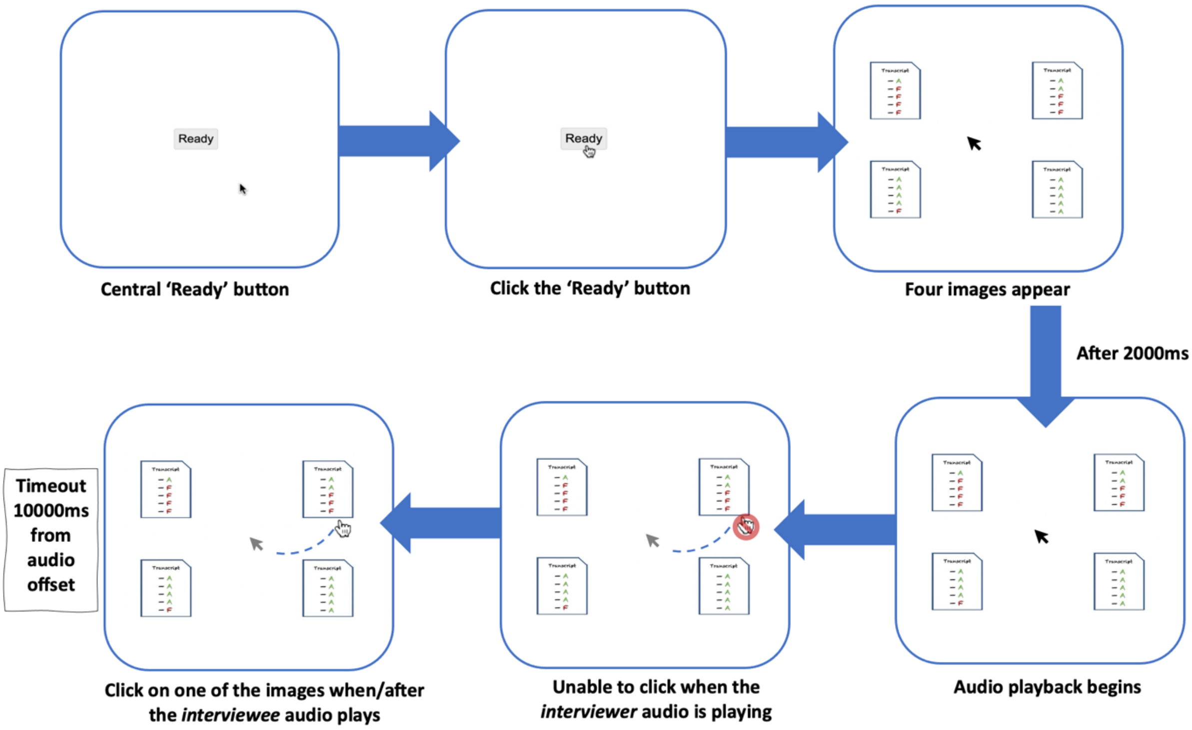


Figure 6.2 Flow of each trial in the mouse-tracking experiment.

Two attention-check trials were included in the middle (after the third interview conversation) and at the end (after the last interview) of each participant's session. In these trials the interviewer gave straightforward information indicating which image should be clicked. For example, in the middle attention check trial, the interviewer said, "Among the candidates we have interviewed so far, I'm thinking about ruling out the one who knows the least about our lab" such that the interviewer question uniquely determined the answer (left-bottom image in Figure 6.3).

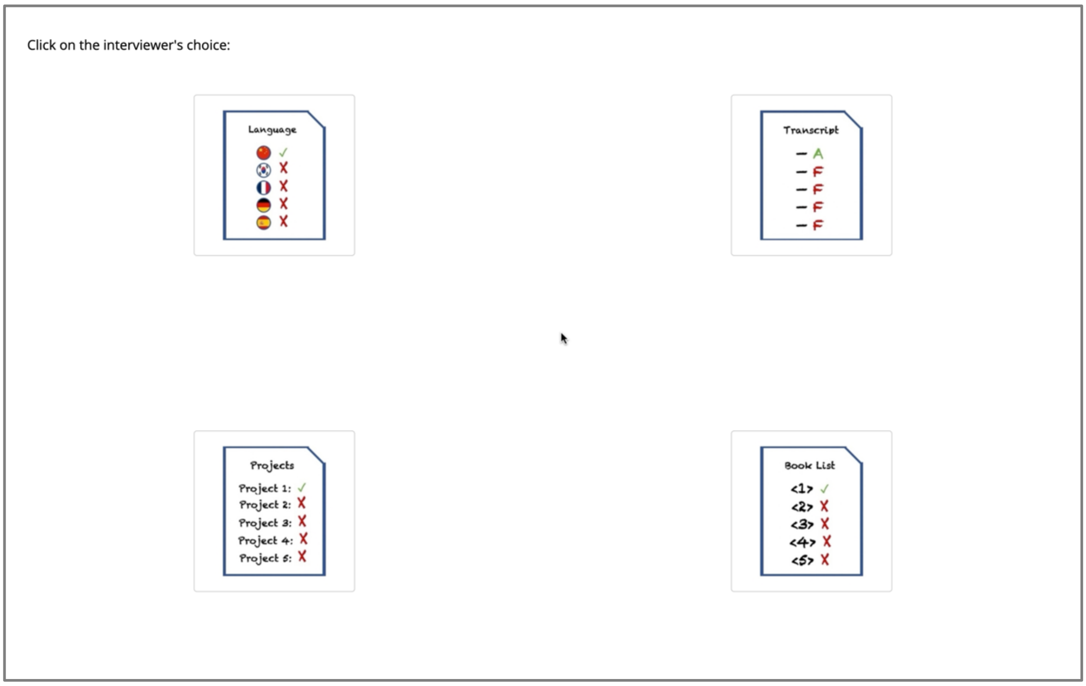


Figure 6.3 Attention-check trial example (mid-attention-check).

The mouse-tracking stage finished with a second attention-check trial, after which participants were directed to a Qualtrics questionnaire page, where they answered questions about the experiment and were asked if they had noticed what the experiment was about.

For each experimental item, we recorded the mouse pointer position over the time as well as the identity of the object clicked.

6.2.2 Results

Data and analysis scripts are available at https://osf.io/m59u3. Prior to analysis, data from filler trials, as well as practice and attention-check trials, were removed. All the statistical analyses were carried out in R version 4.2.3 (R core team, 2023). In the following analysis, an effect with p-value less than 0.05 or absolute t value larger than 2 (following Baayen, 2008) is regarded to be significant.

6.2.2.1 Final object click

Click results

Total numbers of clicks on each image by condition (disfluent/fluent), as well as the percentages of participants clicking on each image in each condition, are shown in a bar plot presented in Figure 6.4.

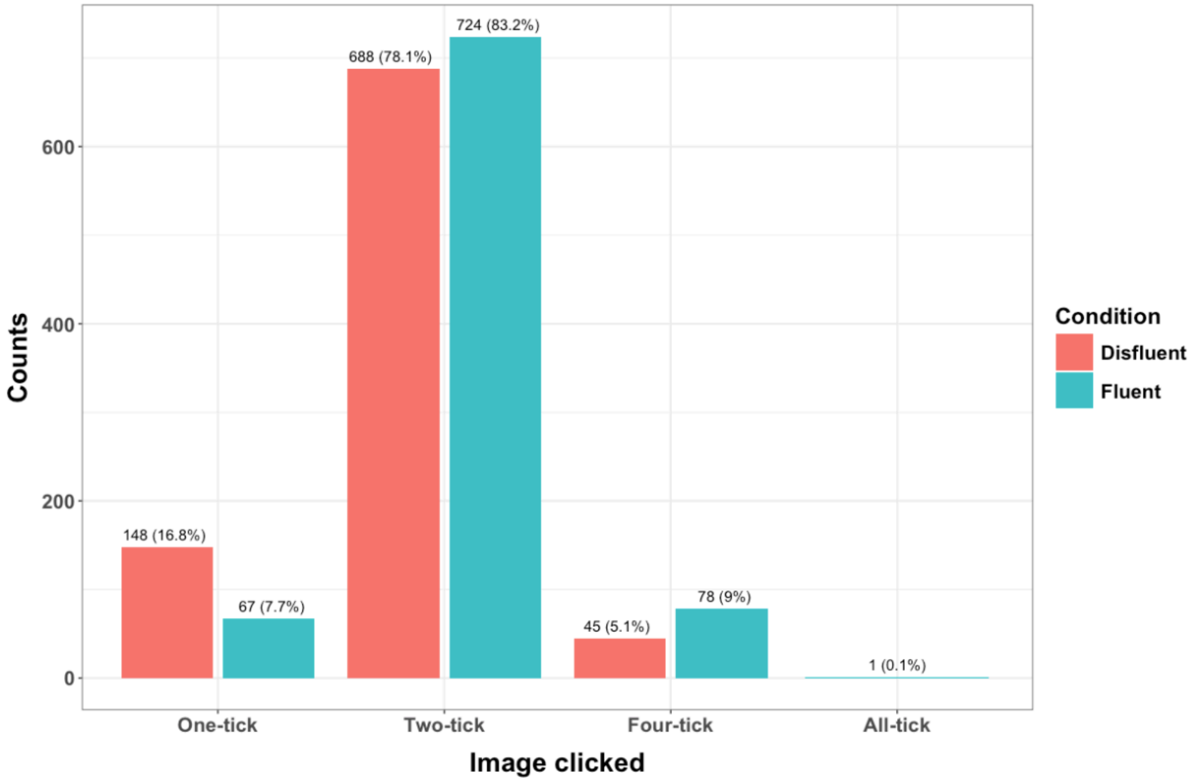


Figure 6.4 Bar plot showing total number and distribution of mouse clicks recorded on each image (One-tick, Two-tick, Four-tick, or All-tick) by manner of delivery (disfluent/fluent).

For the disfluent compared to the fluent conditions, there were fewer clicks on two- and four-tick images, but more clicks on the one-tick image. Clicks on the one-tick image were modelled using a mixed-effects logistic regression with one fixed effect for the within-participant and within-item predictor, the manner of delivery (fluent vs. disfluent), and random effects with by-participant and by-item random intercepts and slopes. Participants were more likely to click on the one-tick image following a disfluent utterance compared to a fluent utterance ($\beta = 8.074$, $SE = 0.591$, $p < 0.001$). This response pattern is in keeping with the predictions for the social reasoning account, suggesting that the presence of disfluency does indeed bias interpretation in favour of the socially undesirable meaning, here the pragmatic 'some but not all' interpretation of *some*.

Time taken to click

After hearing the critical word *some*, the time participants took to click their final pick of image is shown in Figure 6.5. We grouped participants' choices into two groups: clicks on the one-tick image and clicks not on the one-tick image.

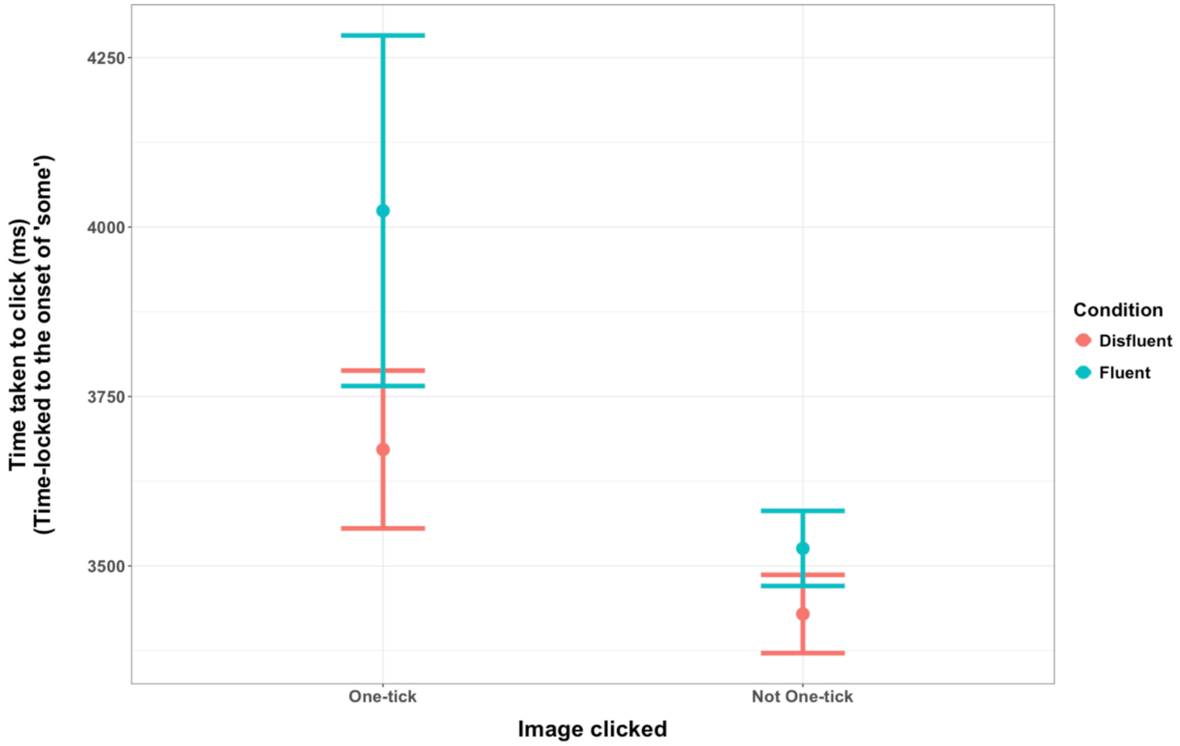


Figure 6.5 Average time to click (relative to the onset of critical word *some*) on each group of choices (clicks on the one-tick image; clicks not on the one-tick image) by manner of delivery (disfluent/fluent). Error bars represent ±1 SE of the mean clicking time.

The overall time participants spent in deciding which image to click, relative to the onset of *some*[17], was modelled using a linear mixed effects regression. The model included the fixed effects for the manner of speech (fluent vs. disfluent), whether the participant clicked on a one-tick image, and their interaction. Additionally, the model accounted for random effects with by-participant random slopes for clicks on one-tick

[17] We calculated overall reaction time relative to the onset of *some* here to avoid the length feature of different audios influencing the statistical results. As some items' audios were originally longer than the others, therefore participants' reaction time hearing those audios was expected to be longer in nature. Time-locking the overall time to the onset of some was to eliminate this influence.

images and random intercepts, as well as by-item random slopes for disfluency and random intercepts. For participants who eventually clicked on the one-tick image, they took significantly longer to decide and click after hearing *some* in fluent condition compared to in disfluent condition ($\beta = 531$, $SE = 228$, $t = 2.33$). This pattern of response time, coupled with participants' clicking patterns, is also keeping in with the predictions of social reasoning account, suggesting that the presence of disfluency biases a socially undesirable interpretation, affecting both the choice of the image clicked and the time taken to make the decision.

6.2.2.2 Mouse movements

The mouse-tracking data obtained from all 150 participants were processed prior to conducting the analysis. For each trial for each individual, mouse trajectories were time-normalised into 101 time-steps (following Dale, Kehoe, & Spivey, 2007 and Spivey, Grosjean, & Knoblich, 2005). Step 0 represents the onset of the interviewee audio and step 100 corresponds to the time point when the participant made the click. All movements were calculated relative to an origin (0,0) which represented the centre of the screen.

Participants' mouse trajectories towards each image in each condition (disfluent/fluent) are shown in Figure 6.6. Each line represents the average trajectory of the mouse pointer, starting at the onset of the interviewee's response and ending when an image is clicked. Coloured points indicate relative time. The colours (from red points at the centre of the screen to violet in each corner) indicate each 10% of trajectory time.

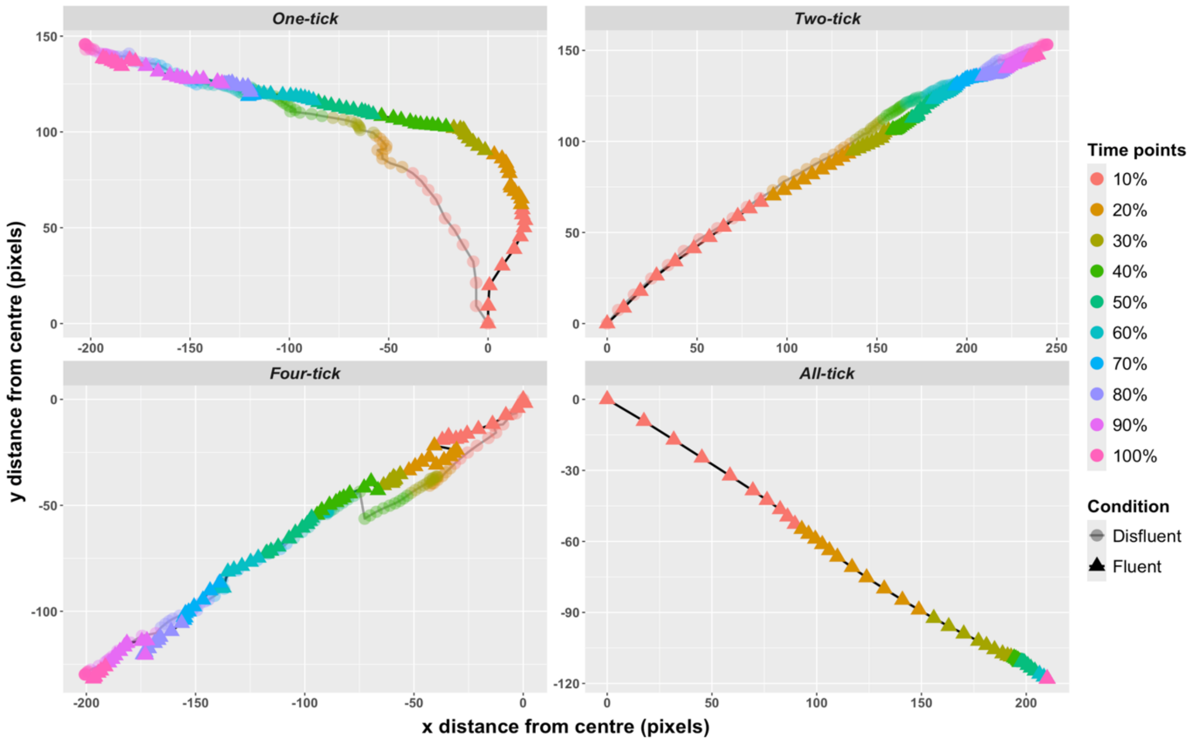


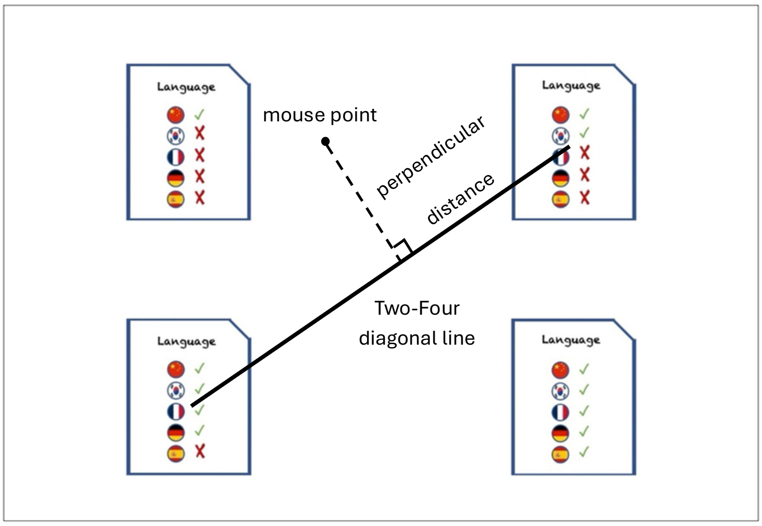
Figure 6.6 Aggregated mouse trajectories towards four images (one-, two-, four-, all-tick) by condition (disfluent/fluent), and the colours of points (from red points at the centre of the screen to violet in each corner) indicate 10%, 20%, 30%...100% of trial time. Note that the disfluent condition is represented with the grey line with round points in relatively lower saturation, and the fluent condition is represented with the black line with triangle points in normal saturation.

The all-tick image was only selected in one trial. As can be seen in Figure 6.6, the pointer trajectories for two- and four-tick images show similar patterns in each condition, both in terms of path and in terms of distance over time. However, the patterns of mouse movements in which participants eventually click on the one-tick image show obvious differences between conditions. Participants appear to move faster and more directly towards the one-tick image in the disfluent condition. In the fluent condition, they make hesitant movements towards other images, taking more of the period between the onset of the interviewee's response and their own mouse click to make a decision to move towards, and eventually click, the one-tick target.

6.2.2.3 Mouse trajectory analysis

To further explore differences in pointer trajectories between disfluent and fluent conditions, we used the Bootstrapped Differences of Timeseries (BDOTS) package (version 1.2.5, Nolte, Seedorff, Oleson, Brown, Cavanaugh, & McMurray, 2023) in R to analyse our mouse-tracking data. BDOTS is a statistical method which is designed to estimate the time window during which two time curves differ (see Oleson, Cavanaugh, McMurray, & Brown, 2017; Seedorff, Oleson, & McMurray, 2018 for details). Its ability to detect subtle differences in the temporal dynamics of mouse movements makes it suitable for investigating real-time processing.

To perform this analysis, we created a distance variable, which represents the perpendicular distance from each mouse point to the diagonal line formed by joining the centres of the two- and four-tick images, henceforth referred to as the two-four diagonal line (see Figure 6.7). We chose this diagonal line as a reference line because the two- and four-tick images both represent meanings that are compatible with the pragmatic (some-but-not-all) meaning of *some* and are the most clicked choices. The perpendicular distance from this line, therefore, depicts the extent to which a participant moves away from common, plural, understandings of *some*. In this sense, analysing this distance over time allows us to map listeners' **non-plural tendencies** when hearing disfluent/fluent utterances.[18]



[18] We appreciate that movements towards items at the bottom right (all-tick) represent a plural interpretation of *some* (as "all"). However, since all-tick was only selected in one instance, "non-plural" seems to be the most useful term to use for expositional purposes.

Figure 6.7 A visual representation of the perpendicular distance from a given mouse point to the diagonal line formed by two- and four-tick image.

The processed data were fit using a logistic function to compare the difference of curves for disfluent and fluent conditions (Top panel in Figure 6.8), regardless of where the final clicks are. In the fitting stage, 6 curves had good fits with AR1 (R^2 >= .95), 22 curves had good fits without AR1, 3 cusrves had reasonable fits with AR1 (R^2 >= .80) and 58 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.994 and the adjusted alpha was calculated to be 0.00931. We found regions of significance starting from time step 27 to the end of the trial time (Bottom panel in Figure 6.8).

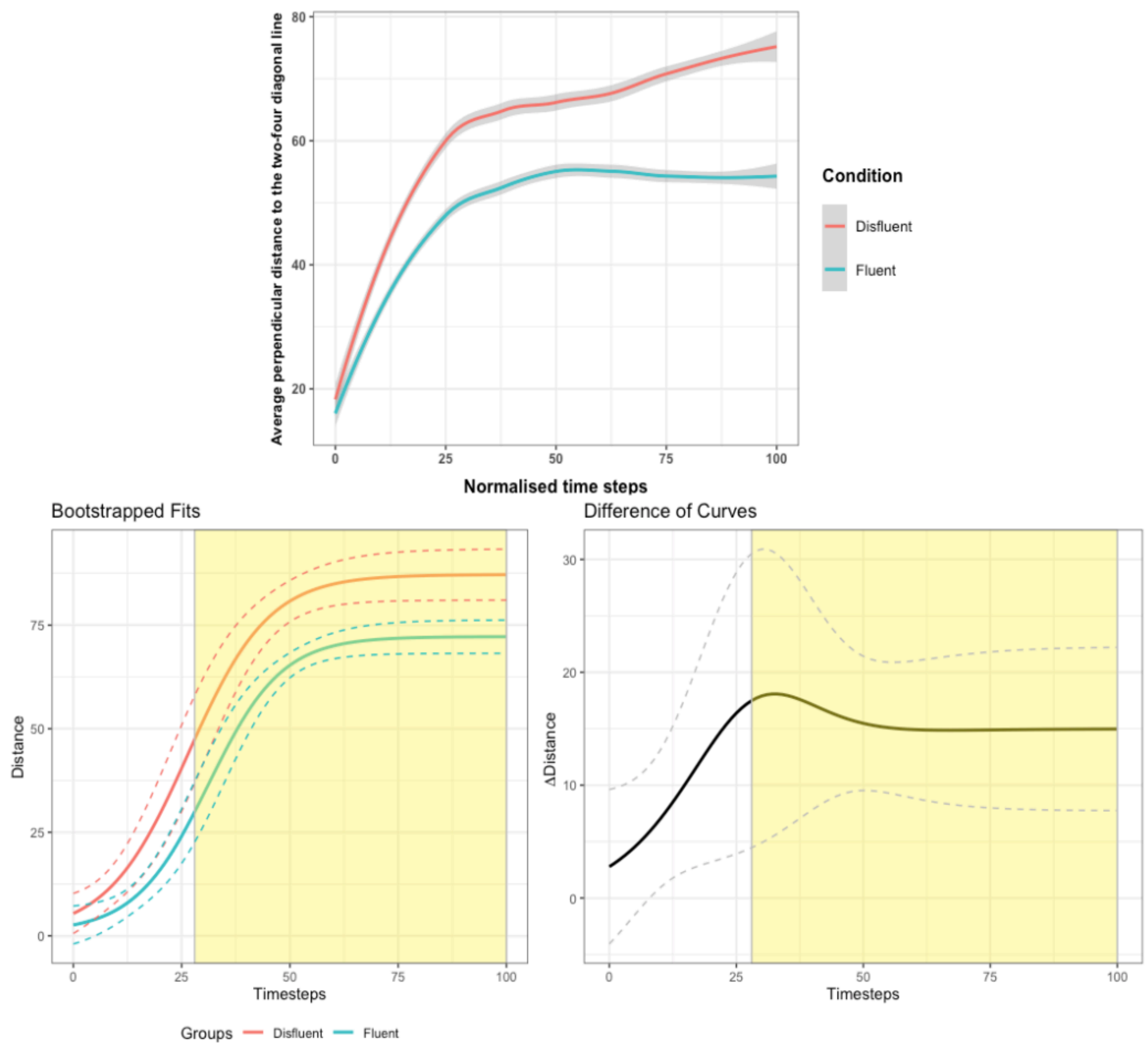


Figure 6.8 *(Top panel)* Two time-curves showing average perpendicular distance (in pixels) from the mouse pointer to the two-four diagonal line across 101 time steps by disfluency conditions; *(Bottom panel)* BDOTS results of curve differences between conditions (disfluent/fluent) highlighting the significant time period of differences in Experiment 3.

The BDOTS analysis shows that participants were faster to move away (in perpendicular distance) from the two-four diagonal line in the disfluent compared to the fluent condition. It indicates that following disfluent utterances, listeners are more likely to select a non-plural meaning of *some* (higher asymptote) and to make that decision quickly (steeper slope). In particular, differences between conditions emerge at step 27 in the mouse trajectory.

Looking at the disfluent condition in isolation, we further explored the mouse trajectories dependent on where participants eventually clicked. Data in the disfluent condition were fit using a logistic function to compare the distance metrics between those clicking on the one-tick image (noted as **singular** image) and those clicking on other images (**plural** images) (Top panel in Figure 6.9).

In the fitting stage, 11 curves had good fits with AR1 ($R^2 >= .95$), 34 curves had good fits without AR1, 5 curves had reasonable fits with AR1 ($R^2 >= .80$) and 87 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.997 and the adjusted alpha was calculated to be 0.0139. We found regions of significance starting from time step 36 to the end of the trial time (Bottom panel in Figure 6.9).

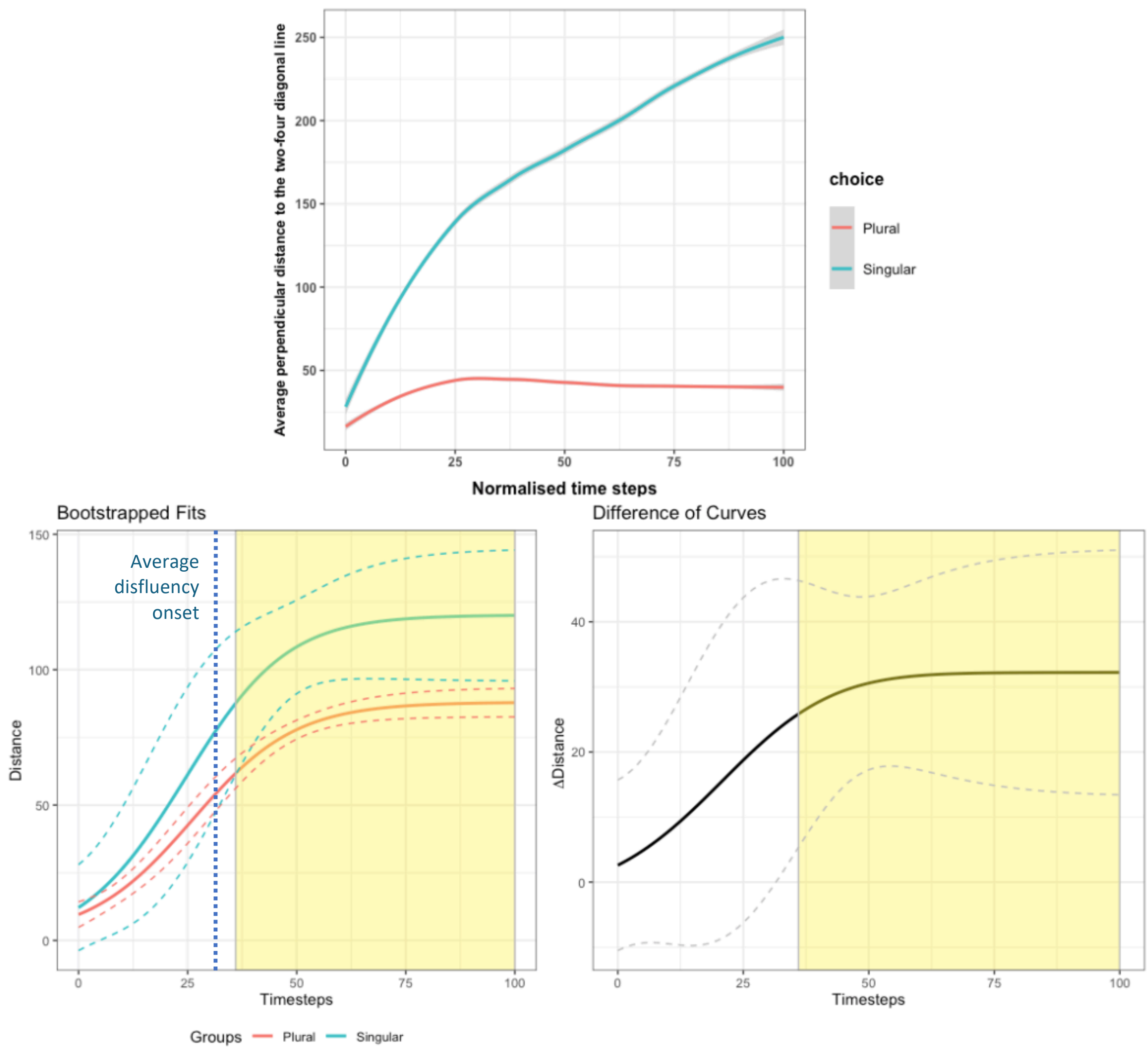


Figure 6.9 *(Top panel)* Two time-curves showing average perpendicular distance from the mouse pointer to the two-four diagonal line across 101 time steps by where the final click is singular (one-tick) or plural (two-, four-, all-tick) images; *(Bottom panel)* BDOTS results of curve differences between final clicks on singular option (one-tick image) and plural options (two-, four-, all-tick images) with the significant time period of differences highlighted and average disfluency onset noted (at around step 30) in Experiment 3.

Mapping the results within the disfluent condition on to actual timings of mouse movements shows that the non-plural tendency emerged quite quickly: When participants encountering disfluency decided to click the one-tick image, their mouse movements diverged around 355ms after average disfluency onset. This indicates that individuals decided to click on the one-tick image almost as soon as they heard the disfluency.

6.2.3 Discussion

Experiment 3 builds on prior work which has established an effect of disfluency on the interpretation of *some*, but we introduce an alternative context for which the heightened attention and social reasoning accounts make different predictions. Specifically, we tested whether disfluency preceding *some* caused listeners to pay increased attention to the literal meaning of ‘some and possibly all’, or whether listeners tended to interpret *some* in a social context, with disfluency heightening the salience of a socially undesirable interpretation.

Experiment 3 supports the social reasoning account in two ways. First, in a context where it is desirable to have more qualifications, disfluency biases towards socially undesirable pragmatic interpretations of *some* which imply smaller numbers. In fact, it would appear that the meaning of *some* is stretched, in that participants accepted an interpretation where *some* was associated with the meaning ‘one’ in 17% of disfluent and 8% of fluent trials. Taken together with the findings from Loy et al. (2019), this suggests that listeners take the social context into account when reasoning about scalar quantifiers such as *some*. Second, this reasoning happens very quickly: Where utterances are disfluent, listeners who are going to choose the one-tick image make the decision to select that target within 355ms, and initiate their mouse movements appropriately. Where the utterances are fluent, they are more hesitant, making early mouse movements which suggest that the two-tick interpretation of *some* is initially in contention.

However, even though the value of one is compatible with *some* (in its ‘at least one’ meaning), interpreting *some* as meaning one is still an unexpected finding. This leaves open an alternative interpretation, which Experiment 4 aims to explore.

6.3 Experiment 4

Experiment 3 shows that social reasoning, together with manner of delivery, can affect the ways in which listeners interpret the meaning of *some*. However, the surprising finding that clicks on one-tick targets became more frequent following disfluent utterances raises the question of how that meaning emerges. On the one hand, comprehenders could be using a *stretching* strategy: Disfluency in context causes them to stretch the meaning of *some* to the lowest possible boundary compatible with the meaning of *some* ('at least one'). On the other hand, listeners might decide that a disfluency can *overwrite* the meaning of *some* in the right context, perhaps because disfluent filled pauses, specifically *um* and *uh*, have been shown to act as cues to deception (DePaulo, Rosenthal, Rosenkrantz, & Green, 1982; Loy, Rohde, & Corley, 2016). If the listener infers that the speaker is being deceptive, they may be able to ignore the meaning of *some* entirely.

To distinguish these two possibilities, Experiment 4 replicated Experiment 3 with one change to the design. In Experiment 4, the one-tick images were replaced with zero-tick images. The number zero is not compatible with any meaning of *some*. If listeners are still making interpretations based on the meaning of *some,* as supposed by the *stretching account*, we would not expect to see clicks on the zero-tick image. However, if listeners are *overwriting some* when speakers in this experiment are disfluent, they may click on the zero-tick image.

6.3.1 Methods

6.3.1.1 Participants

We recruited participants and filtered collected data in the same way as for Experiment 3. Mouse movement data from one hundred and seventy-three (Female: 87; Male: 82; Non-binary: 4) self-reported native English speakers recruited via Prolific Academic were analysed further, with data from an additional 11 participants ruled out (either because they took more than 40 minutes to finish the whole session, or they failed one or both attention checks). Participants provided informed consent by clicking the "I consent" button on a consent page following the University Research Ethics Committee guidelines (reference number: 84/2021-8).

6.3.1.2 Material and experiment design

As in Experiment 3, we recorded participants' mouse movements in a web-based task in which we manipulated Disfluency (present vs. absent) within-subjects in a set of target trials (N=12). For each target trial, participants saw four images with four different counts of qualifications displayed on the screen (see Figure 6.10 for example trials) and heard recorded interview conversations between an interviewer and one of six interviewees. Experiment 4 was implemented in jsPsych version 7.0 (De Leeuw, 2015).

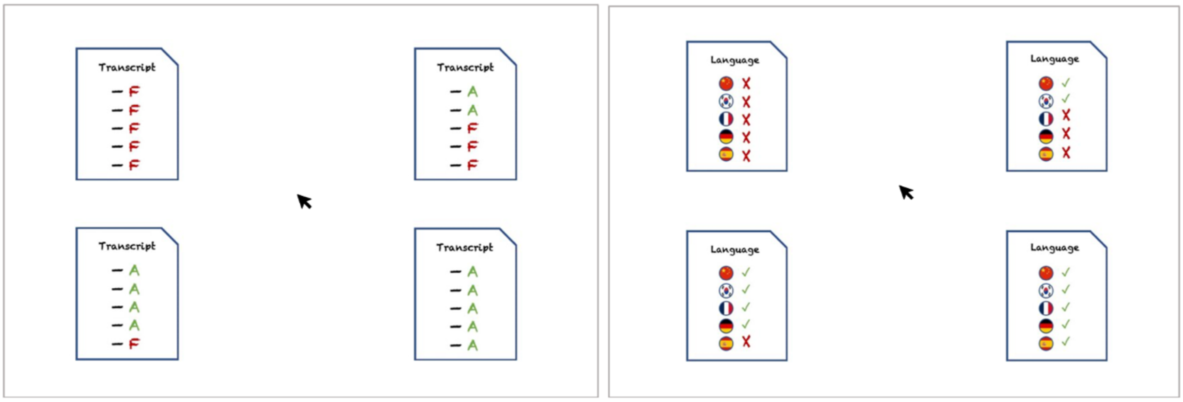


Figure 6.10 Displays of two target trials in Experiment 4, same as Experiment 3 except that the top left in both panels consists of a zero-qualification image.

6.3.1.3 Audio and visual stimuli

The audio stimuli from Experiment 3 were used in Experiment 4. For the visual stimuli, participants saw four images from one set with zero, two, four or all (five) qualifications marked with green ticks and red crosses (green “A” and red “F” in one case). As in Experiment 3, the images always appeared in fixed locations (with zero-tick on the top left, two-tick top right, four-tick bottom left and all-tick bottom right).

6.3.1.4 Procedure

Other than the use of zero-tick in place of one-tick images, the procedure was identical to that for Experiment 3.

6.3.2 Results

We analysed the data from Experiment 4 in the same way as for Experiment 3, in R version 4.2.3 (R core team, 2023). Filler trials, practice trials and attention-check trials were removed prior to analysis.

6.3.2.1 Final object click

Click results

Figure 6.11 shows the total numbers, as well as the percentages, of clicks on each image by condition (disfluent/fluent). Similar to the clicks in Experiment 3, participants' responses suggest a general bias towards a non-plural option cued by disfluency: Following disfluent utterances compared to fluent utterances, people clicked more on the zero-tick image (19% vs. 6%) and less on two- and four-tick images (81% vs. 93%). As in Experiment 3, where clicks on the one-tick image increased by 10% following disfluency, participants' choices are influenced by speakers' manner of speech ($\beta = 11.83$, SE = 0.62, $p < 0.001$). This response pattern is in keeping with the predictions of the *overwriting* account, suggesting that the presence of disfluency allows participants to consider the zero-tick interpretation of the interviewee's utterance, which is incompatible with standard meanings of *some*.

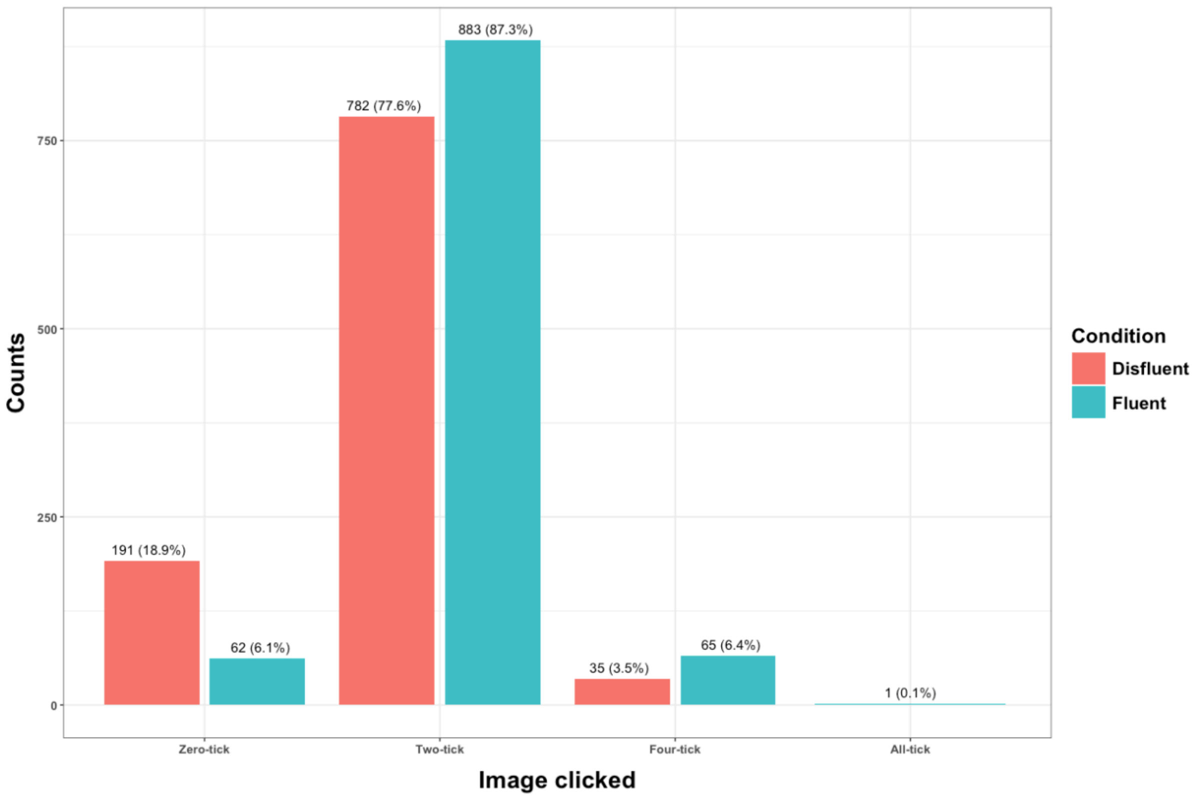


Figure 6.11 Bar plot showing total number and distribution of mouse clicks recorded on each image (Zero-tick, Two-tick, Four-tick, or All-tick) by manner of delivery (disfluent/fluent).

Time taken to click

After hearing the critical word *some*, the time participants took to click their final pick of image is shown in Figure 6.12. As in Experiment 3, we grouped participants' choices into two groups: clicks on the zero-tick image and clicks not on the zero-tick image.

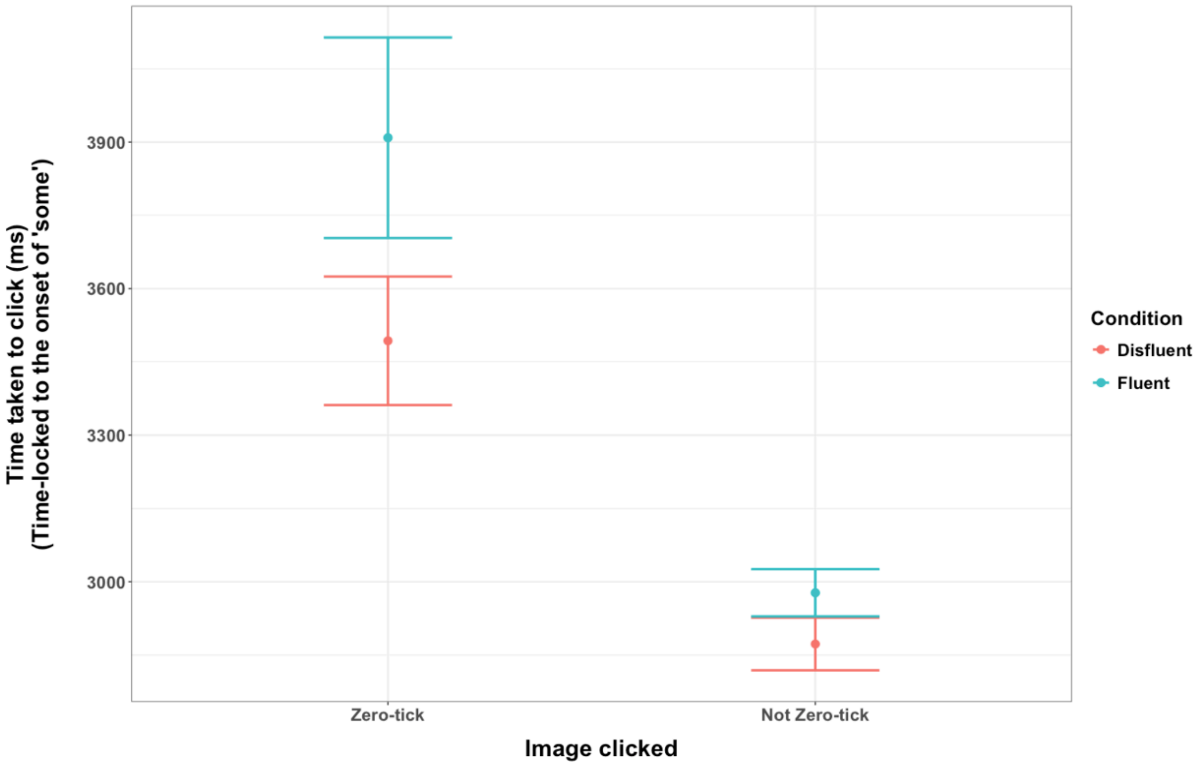


Figure 6.12 Average time to click (relative to the onset of critical word *some*) on each group of choices (clicks on the zero-tick image; clicks not on the zero-tick image) by manner of delivery (disfluent/fluent). Error bars represent ±1 SE of the mean clicking time.

The overall time participants spent deciding which image to click, relative to the onset of *some*, was modelled using a linear mixed effects regression. The model included the fixed effects for the manner of speech (fluent vs. disfluent), whether the participant clicked on a zero-tick image, and their interaction. It also accounted for random effects with by-participant and by-item random slopes for the manner of speech and random intercepts. Consistent with the response time pattern observed in

Experiment 3, participants who eventually clicked on the zero-tick image took significantly longer time to decide and click after hearing *some* in fluent condition compared to the disfluent condition ($\beta = 497$, $SE = 206$, $t = 2.42$). Moreover, following a disfluent utterance, those who did not click on zero-tick images responded significantly faster to response compared to those who eventually clicked on the zero-tick images ($\beta = -471$, $SE = 116$, $t = -4.06$). Importantly, there was a significant interaction effect between the manner of speech and whether participants clicked on a zero-tick image, indicating that in both the disfluent or fluent conditions, not clicking on the zero-tick image reduced response time ($\beta = -428$, $SE = 211$, $t = -2.03$). Despite the clicking results indicating a response pattern consistent with the predictions of the overwriting account, this response time pattern further reveals that participants took longer to decide and click on a zero-tick image compared to other images. The longer response time for zero-tick images encourages a further exploration of participants' detailed mouse movements.

6.3.2.2 Mouse movements

We processed the collected mouse-tracking data from all 173 participants following the same two key steps as in Experiment 3. Although participants appeared willing to interpret *some* as 'zero', inspection of their mouse movements suggest that the processes underlying this decision differ from those which underlie an interpretation of *some* as meaning 'one'. Figure 6.11 shows participants' average mouse trajectories towards each image for fluent and disfluent items. The colours of the points (from red to violet) each indicate 10% of the trial time. We consider the trajectories for fluent versus disfluent conditions in the cases where participants chose either the one-tick interpretation (Experiment 3) or zero-tick interpretation (Experiment 4). When ultimately clicking on the one-tick image (top-left panel in Experiment 3, see Figure 6.5), participants show hesitant and less direct movements when the utterance is fluent. However, when ultimately clicking on a zero-tick image (top-left panel in Experiment 4, see Figure 6.13), participants show hesitant movement towards other images in both conditions.

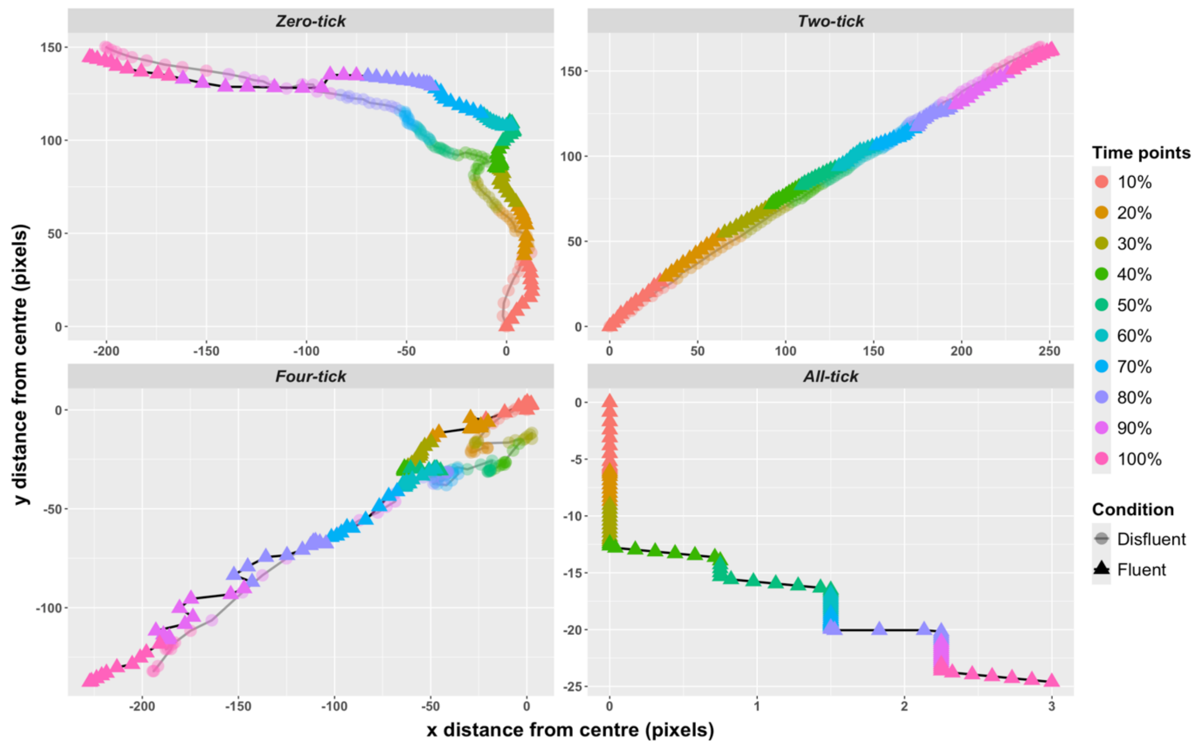


Figure 6.13 Aggregated mouse trajectories towards four images (zero-, two-, four-, all-tick) by condition (disfluent/fluent), and the colours of points (from red points at the centre of the screen to violet in each corner) indicate 10%, 20%, 30%...100% of trial time. Note that the disfluent condition is represented with the grey line with round points in relatively lower saturation, and the fluent condition is represented with the black line with triangle points in normal saturation.

6.3.2.3 Mouse trajectories analysis

Again, we calculated the perpendicular distance between mouse pointers to the two-four reference line to indicate the non-plural tendency. Obtained values were fit using a logistic function in a BDOTS analysis. We first compared the difference of time-curves for disfluent and fluent conditions across all items (Top panel in Figure 6.14).

In the fitting stage, 12 curves had good fits with AR1 (R^2 >= .95), 26 curves had good fits without AR1, 16 curves had reasonable fits with AR1 (R^2 >= .80) and 61 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.999 and the adjusted alpha was calculated to be 0.0274, showing that we failed to find any significant regions where these two time-curves start to differ. This result (Bottom panel in Figure 6.14) suggests that, in this experiment where zero-tick replaces one-tick image, manner of speech does not significantly influence the timing of participants' choice towards which image to click on.

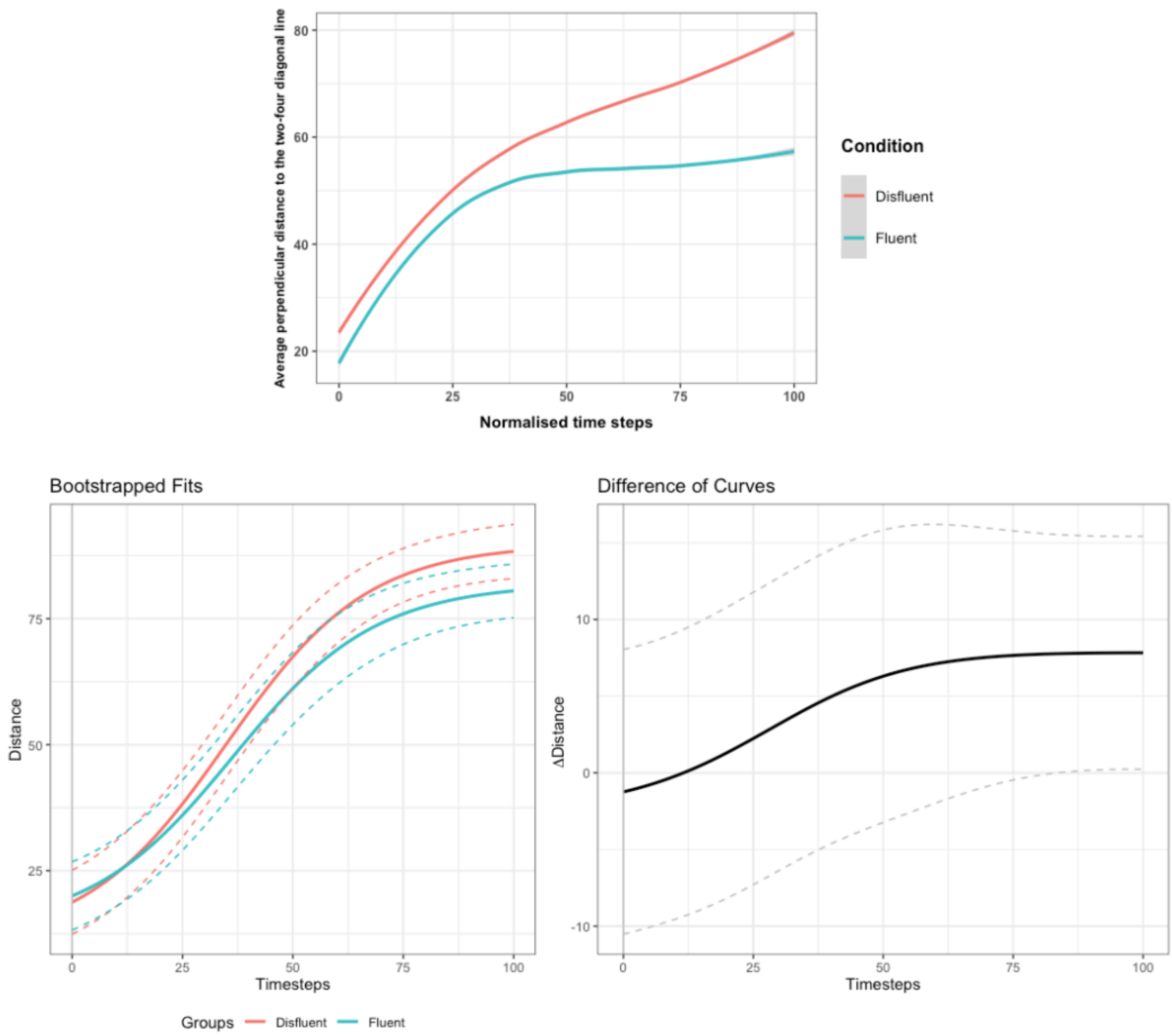


Figure 6.14 *(Top panel)* Two time-curves showing average perpendicular distance (in pixels) from the mouse pointer to the two-four diagonal line across 101 time steps by disfluency conditions; *(Bottom panel)* BDOTS results of curve differences between conditions (disfluent/fluent) highlighting the significant time period of differences in Experiment 4.

As for Experiment 3, we performed a second analysis, focusing on the disfluent condition. We fit the data using a logistic function, to compare the curve differences between participants clicking on the zero and those clicking on plural images (Top panel in Figure 6.15).

In the fitting stage, 25 curves had good fits with AR1 (R^2 >= .95), 42 curves had good fits without AR1, 20 curves had reasonable fits with AR1 (R^2 >= .80) and 90 curves had reasonable fits without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.996 and the adjusted alpha was calculated to be 0.0114. We found regions of significance starting from time step 64 to the end of the trial time (Bottom panel in Figure 6.15).

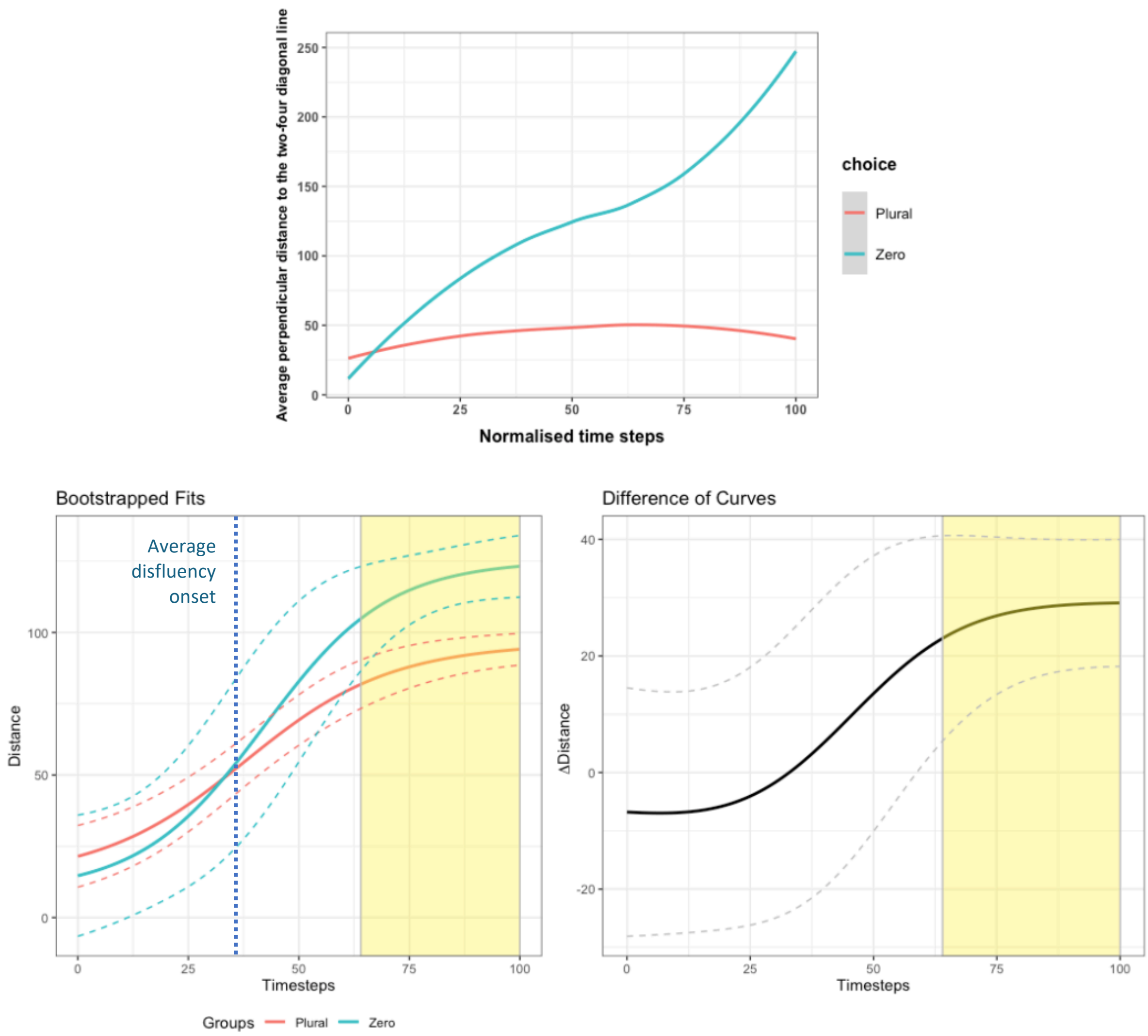


Figure 6.15 *(Top panel)* Two time-curves showing average perpendicular distance from the mouse pointer to the two-four diagonal line across 101 time steps by where the final click is zero (zero-tick) or plural (two-, four-, all-tick) images; *(Bottom panel)* BDOTS results of curve differences between final clicks on zero option (zero-tick image) and plural options (two-, four-, all-tick images) with the significant time period of differences highlighted and average disfluency onset noted (at around step 36).

When encountering disfluent utterances, the mouse movements in the trials where participants clicking on the zero-tick image diverged at around step 36 from those in the trials where other images are clicked. If mapping this time onto actual mouse movement timings, this indicates that participants decide to choose the zero-tick image over the other ones around 1357ms after average disfluency onset, a delay compared to the decision point in Experiment 3 (335ms) which suggests that overwriting *some* to mean 'zero' may be a slower process than stretching it to mean 'one'.

6.4 General Discussion

The current study explores how listeners' interpretations of the ambiguous scalar quantifier *some* are influenced by the speaker's manner of speech. Different from the context used in Loy et al. (2019) where larger values are socially undesirable ("I ate, uh, some oreos"), we set up a social context where the undesirable meaning corresponds to smaller values of *some* ("I've got, uh, some 'A's"). The mouse-tracking results that we report here show that disfluency has the opposite effect from that shown in Loy et al.'s study, reducing the values participants associate with *some*: We found that participants are more likely to select images corresponding to one or zero qualifications, following disfluent utterances. These results complement the findings of Loy et al. (2019), establishing that both results likely stem from participants' social reasoning about the presence of disfluency in a context with *some*.

Moreover, differences in participants' mouse movement patterns show that given the same social context, comprehenders are relatively quick to commit to one qualification (Experiment 3) and slow to commit to zero qualifications (Experiment 4). This result not only confirms that social context and manner of speech can combine to affect the interpretation of *some*, but also suggests that extending the meaning of *some* to 'one' is relatively easy, but overwriting it with 'zero' (in effect, deciding that a speaker is lying) may be more demanding.

A potential account of our findings is that both the stretching and overwriting strategies are in play when listeners are searching for an interpretation of ambiguous *some*. A context such as that of an interview may set up an expectation that speakers could potentially be misleading or deceiving. Bearing possible deception in mind, listeners who encounter disfluency try to push the boundary of the meaning of *some* as low as possible. When the lowest number available is still within the boundary of the semantic meaning of *some* ('one'), listeners seem to arrive at this associated meaning relatively quickly. However, when the available lowest number is not included in the meaning of *some* ('zero'), listeners hesitate before overwriting the interpretation to zero, if they are going to choose that option.

Critically, our findings emphasise the necessity of looking at the lower bound of the meaning of *some* ('one') to truly differentiate whether it is the context or other factors–influencing listeners' interpretation of *some*. Even though the results of Bonnefon et al. (2015) and Loy et al. (2019) hint at a potential role of social context in interpreting *some*, these two prior studies focus on the upper bound interpretation ('all') and fail to rule out the possibility of heightened attention cued by disfluency. With a context pushing people's interpretation to the lower bound meaning of *some*, the current study manages to distinguish the possible accounts and confirms the social role of disfluency in interpreting *some*.

For both one- and zero- qualification images, there are similarly increasing numbers of clicks following disfluent utterances. However, analysis of the real-time mouse movement patterns reveals that listeners do not reach these choices via the same processes. Instead, in finding evidence that similar behavioural outcomes might be achieved by distinct cognitive processes (Hala, Pexman, & Glenwright, 2007; Happé, 1995; Jolliffe & Baron-Cohen, 2000), we highlight the importance of exploring the timecourse of comprehension.

Experiment 4 demonstrates that, in relevant circumstances, listeners may choose to overwrite the meaning of *some,* although at some cognitive cost relative to extending its meaning in Experiment 3. This raises a question concerning studies which have varied the salience of *all*, the upper-bound semantic meaning of *some* (e.g., Breheny, Ferguson, & Katsos, 2013; Breheny et al., 2006; Politzer-Ahles & Husband, 2018). Although *all* is compatible with a semantic interpretation of *some*, it is also feasible that a pragmatic interpretation is simply overwritten (perhaps with literal *all*) in relevant circumstances. Explorations of the upper-bound are unlikely to distinguish these accounts (especially without information on the timecourse of interpretation). Now that the possibility of overwriting has been raised, and demonstrated, it would appear that investigating lower-bound implicatures of *some* may be more profitable.

Methodologically, our current study shows the possibility of using a web-based mouse-tracking task to explore real-time processing, with the advantage of easy implementation and quick recruitment of participants. With four images on the screen, our materials were more complex than those in previous mouse-tracking studies (with two images located symmetrically). As such, our study confirms the viability of studying more complex visual scenes and of using visualisation and time-course analysis of trajectories that go beyond standard methods of analysis for mouse-tracking data (Spivey, Grosjean, & Knoblich, 2005).

Our two experiments suggest that manner of speech facilitates listeners' understanding of the scalar quantifier *some* via a process of social reasoning about speakers' goals, and that this reasoning happens fast, in the early stages of listeners' comprehension. Given the role of social reasoning, these findings open up new questions about possible differences in this kind of reasoning across individuals with varying pragmatic abilities or across contexts that evoke different conversational goals or speaker agendas.

6.5 Conclusions

This chapter reported two additional mouse-tracking experiments which further investigated the pragmatic effect of disfluency in listeners' real-time processing of ambiguous scalar quantifier *some*. When varying the social context to one where the smaller values are socially undesirable, disfluency showed an opposite effect found in the experiments in Chapter 5, reducing the values listeners associated with *some* to "one" or even "zero". The current results are in keeping with an account in which the effects of disfluency reflect social reasoning and contrary to one in which such effects depend on heightened attention. Moreover, listeners appeared to be quick to commit to one qualification (Experiment 3) and slow to commit to zero (Experiment 4), suggesting that social context and manner of speech can combine to affect the interpretation of *some* as an utterance unfolds. Extending its meaning to *one* is relatively easy, but overwriting it with zero – in effect, deciding that a speaker is lying – is more demanding. Chapter 7 further tests whether listeners' pragmatic reasoning of disfluent utterances varies systematically across individuals.

Chapter 7
Disfluency, Deception, and Individual differences

Findings in the mouse-tracking experiments presented in Chapter 5-6 highlight that listeners engage in a social reasoning process cued by disfluency in interpreting ambiguous word *some*. These results raise the question of whether listeners' pragmatic abilities influence the effect of disfluency in comprehending utterances. In this chapter, I present an eye-tracking experiment, testing whether the reasoning process varies across individuals with varying levels of pragmatic ability suggested by Autism Quotient scores.

The experimental work presented in this chapter are reported by Li, Rohde, & Corley (2020). The publication is reproduced in Appendix A.

7.1 Introduction

Even though Autism Spectrum Disorder (ASD) is generally associated with a deficit in social and figurative aspects of communication, many previous studies investigating the abilities of participants with ASD to understand non-literal language have shown that these participants obtain similar comprehension scores to their neurotypical peers (e.g., Happé, 1994, 1995; Hala, Pexman, & Glenwright, 2007; Jolliffe & Baron-Cohen, 2000; Loukusa & Molianen, 2009). This raises the question of whether individuals with ASD and those without achieve similar outcomes by employing different processes. Here, we develop the theme of processing differences, with a focus on how the manner of a spoken message is delivered, in particular, disfluent fillers such as "um" and "uh", influences individuals' comprehension.

Fillers are often regarded as socially oriented, influencing discourse dynamics by controlling the flow of conversation (e.g., Clark & Fox Tree, 2002), signalling hedges in conversation (Smith & Clark, 1993), or being taken as cues to deception (Loy, Rohde, & Corley, 2016; Vrij & Semin, 1996; Zuckerman, Koestner, & Driver, 1981). Previous studies have shown that high-functioning individuals with ASD produce fewer fillers than do the typical population (Lake, Humphreys, & Cardy, 2011). Children with autism produce fewer fillers compared to their typically developing peers (Gorman et al., 2016; Heeman, Lunsford, Selfridge, Black, & Van Santen, 2010; Irvine, Eigsti, & Fein, 2016). This reduced use of fillers, considered as a conversational cue, is consistent with a characterization of speakers with ASD as being less able to take their listeners' perspectives (Colle, Baron-Cohen, Wheelwright, & Van Der Lely, 2008; Paul, Orlovski, Marcinko, & Volkmar, 2009). A decreased attention to social cues may yield difficulty in understanding the implicit social meanings associated with fillers (Irvine et al., 2016).

However, this difficulty may not manifest in comprehension outcomes. Participants who struggle with social cues may nonetheless find ways to interpret fillers in an equivalent way to their peers. By focusing on the outcome of comprehension, usually operationalised as whether the eventual interpretation of an utterance aligns to a norm, it is possible that differences in the processes which underlie these outcomes may be overlooked (McKenna, Glass, Rajendran, & Corley, 2015). McKenna et al. (2015) demonstrated a difference in the processing of a type of figurative language, metonymy, in which an attribute of a concept instantiates that concept (such as the use of Dickens when referring to the books written by Charles Dickens rather than to the writer himself). In their study, participants without diagnoses completed an Autism Spectrum Quotient questionnaire (AQ: Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) and read sentences containing target nouns (e.g., Vietnam, Finland), in contexts supporting literal or figurative, metonymic, interpretations (e.g., During my trip, I hitchhiked around Vietnam/Finland; A lot of Americans protested during Vietnam/Finland). Their results showed an "unfamiliar metonym disadvantage" where participants were slower to read a novel metonym (e.g., Finland) in its figurative sense, compared to a metonym which has its figurative meaning established in daily use (e.g., Vietnam referring to the conflict during the 1950s-1970s). Of relevance to the present study, they found that participants with higher scores on the Autism Quotient measure were relatively slowed down by these novel metonymies, while those with lower numbers of self-reported autistic traits were not. Note that all participants scored 100% on comprehension scores: Differences related to AQ could not be found in any measure of interpretation, but were instead only evident when the time course of the processing of the stimuli was investigated.

Although understanding figurative meanings in written language, such as that investigated by McKenna et al. (2015), can be difficult, understanding non-literal meaning in speech adds additional complexity, because artefacts such as fillers can modify literal meaning. Loy et al. (2017) investigated whether interpretations of a sentence (truthful or deceptive) made by listeners in the general population could be influenced by disfluency. Their experiments used a lie-detection game setting, where participants were presented with two images (named image/distractor image) in every trial and were asked to click on the image that they thought concealed some treasure. For each trial, participants heard a pre-recorded fluent or disfluent utterance describing the location of the treasure. Participants' image clicks revealed a bias to interpret disfluent speech as a cue to speaker deception about the location of the treasure. Importantly, eye movements were also measured, and disfluent speech resulted in more frequent early eye movements towards the distractor item, showing the influence of an early process associating disfluency with deception.

Loy et al.'s (2017) study provides two essential points of reference for the current study. First, the method allows us to directly observe the time-course of disfluency comprehension, providing a way to investigate how the manner in which something is said might affect its ongoing interpretation. Second, along with other situations which depend on understanding the interlocutor's intentionality, the recognition of deception has been shown to be difficult for individuals with ASD: For example, it is harder for children with autism to lie to others and to detect when they are being lied to (Ranick, Persicke, Tarbox, & Kornack, 2013). By using a lie detection paradigm in which disfluency can serve as a potential pragmatic cue, we can examine in detail any differences in the processes of comprehension that might arise between individuals. Specifically, we can measure both participants' eventual interpretations – truth or deception? – and the real-time processes that precede that outcome.

Our experiment replicates Experiment 2 from Loy et al. (2017), with a set of participants from the general population who vary in their Autism Spectrum Quotients (AQ). AQ is not a diagnostic test of ASD, but it is used to assess broader autism phenotype (BAP) metrics for understanding subclinical ASD traits within individuals in the general population who are presumed to have typical developmental profiles. The test comprises 50 questions focusing on five BAPs: social skills, attention switching, attention to detail, imagination, and communication (Baron-Cohen et al., 2001). AQ scores from low to high quantify where a given participant is situated on a putative continuum from neurotypical to autism diagnosis (Baron- Cohen et al., 2001). While our findings do not allow us to draw direct conclusions about comprehension processes in ASD, the study emphasizes the utility of real-time measures in comparing processing differences among populations that vary on an AQ spectrum.

Our experiment measures the influence of manner of delivery on pragmatic comprehension in two ways. First, we measure the outcome of the comprehension process by recording which object people click on. This gives us a direct indication of whether a participant has interpreted a given utterance as truthful, in which case they are likely to click on the named (target) object, or deceptive, in which case they are likely to click on the other (distractor) image. If the results of Loy et al. (2017) are replicable, we expect to see an effect of the presence of disfluency, such that participants are less likely to click on the target following a disfluent utterance. We further assess the process of comprehension. By measuring participants' eye movements time-locked to the speaker's mention of the target object, we can test the time-course of comprehenders' responses to the speaker's manner of delivery. If the results replicate those of Loy et al., we should see a decrease of bias towards the named object in the disfluent condition. Of critical interest in the present study is whether the outcome or process measures interact with AQ. An interaction at the outcome level (item clicked) would suggest that the levels of broader autism phenotypes measured by the AQ affect the eventual understanding of what is said. Regardless of whether such an interaction is found or not, an interaction at the process level (eye movements) would suggest differences in the processes underlying non-literal interpretations, opening the door to research with formally-diagnosed groups of participants.

7.2 Methods

7.2.1 Participants

Ethical approval was given by the PPLS Research Ethics Committee with reference number 100-1920/1. Sixty-two self-reported native English speakers (Male: 18; Female: 44) took part in the experiment. Participants ranged from 18 to 35 years old, mainly comprising students at the University of Edinburgh recruited from a wide variety of disciplines, including Engineering, Chemistry, Mathematics, Law, and Psychology. Participants were paid £5 for participation. Participants had normal or corrected-to-normal vision, and all used their right hands to control the mouse used in the experiment. Participants provided written consent prior to the beginning of the experiment.

7.2.2 Material

7.2.2.1 Autism Spectrum Quotient

The Autism Spectrum Quotient (AQ: Baron-Cohen et al., 2001) questionnaire comprises 50 items reflecting five different broader autism phenotypes either positively (e.g., “I am fascinated by numbers”, “I am often the last to understand the point of a joke”), or negatively (e.g., “I find social situations easy”). Participants are asked to respond based on a 4-point Likert scale varying from “definitely agree” to “definitely disagree”. Each response represents one point; and a participant’s AQ score is calculated by totalling their responses (taking polarity into account), with a maximum score of 50. Studies to date suggest that individuals diagnosed with autism tend to score 32 or more, compared to individuals without autism who tend to average around 16. However, the AQ is not a diagnostic test, and autistic traits are present in the wider community (see Folstein & Rutter, 1977; Persico & Bourgeron, 2006 for details). Therefore, AQ scores should be interpreted very carefully. Participants in the current study were informed of this beforehand and were told that their AQ scores would not be provided to them.

7.2.2.2 Visual Stimuli

The visual world paradigm from Loy et al. (2017, Experiment 2) was adapted for the current study. Visual stimuli included 130 line-drawings from Rossion and Pourtois (2004), which were grouped into pairs across 65 trials (5 practice trials; 20 critical; 40 fillers). For each pair of drawings in each trial, the drawing that the speaker named as concealing the hidden treasure is referred to as the target picture, and the other one the distractor picture. The two pictures were presented vertically centred on the screen, and horizontally centred at 25% (left-hand picture) and 75% (right-hand) of the screen width. The target picture's position on the left versus the right was counterbalanced across items.

7.2.2.3 Audio Stimuli

The audio files used in the current experiment were taken from Loy et al.'s Experiment 2. Participants were told, as a cover story, that the speaker had taken part in an earlier experiment, in which she had been told to try and mislead her partner about the treasure's location by sometimes lying.

In 20 critical trials, the recordings of the speaker describing the purported locations of the 'treasure' were categorized as either fluent (e.g., "The treasure is behind the [target]") or disfluent (e.g., "The treasure is behind thee, uh, [target]"). To ensure that participants were exposed to the same utterance (bar disfluency) across conditions, and the same disfluency across disfluent trials, complete fluent and disfluent sentences were first recorded, and then a prolonged article followed by a filler cut out from a disfluent utterance was spliced into each relevant fluent utterance to create the corresponding disfluent version. Disfluent utterances were therefore characterized by a prolonged article ("the" pronounced /ðiː/, or "thee") and the disfluent segment (i.e., "uh") before the target noun.

The assignment of fluent/disfluent conditions to items was counterbalanced across two lists. However, due to a coding error, one critical item ("lamp") in List 1, which should have been in the fluent condition, was mistakenly paired with a disfluent recording. Thus, the final numbers of items in experimental conditions for participants using audio files from List 1 were n=9 (fluent items) and 11 (disfluent items).

To obscure the experimental manipulation, and reinforce the impression that participants were listening to natural recordings, forty additional filler trials referred to pairs of images that were not used in critical trials, and were designed such that half included disfluencies other than “uh”, or discourse manipulations such as discourse markers (e.g., “Okay, the treasure is behind the [target]”) and modals (e.g., “The treasure could be behind the [target]”), and the other half were fluent.

7.2.3 Procedure

The experiment was in two parts, a computer-based lie detection game and an Autism- Spectrum Quotient (AQ) measurement. The researcher began by introducing the process of the whole experiment and the lie detection game task. The eye-tracking experiment which followed was presented using OpenSesame version 3.2.7 (Mathôt, Schreij, & Theeuwes, 2012), with movements of the right eye monitored using an Eyelink 1000 Tower Mount system. Participants were seated at a viewing distance of 80cm from a 19” CRT monitor. They read the instructions first which explained that their aim was to collect as much treasure as possible by clicking on the object concealing the treasure on each trial.

Following reading the brief instructions, a press of the spacebar started the calibration of the eye-tracker. Five practice trials followed the calibration; at the end of the practice session, participants pressed the space bar to commence the sixty experimental trials (20 critical trials and 40 fillers).

Each new trial began with a black fixation dot at the centre of the screen, and participants were told to carefully stare at the black dot when they were ready to start a trial, allowing a drift correction to be applied if necessary. As soon as the fixation dot disappeared, a pair of images – a target and a distractor – appeared on the screen. After a preview period of 2000ms, the mouse pointer appeared at the centre of the screen, and the audio stimulus started. Participants were instructed to click on the object which they believed the treasure was behind as soon as they could. Once one object had been clicked on, the images disappeared, providing visual feedback that the click had been recorded. 1000ms after the end of the audio, or once the participant had clicked on an object – whichever occurred later – a black dot appeared which indicated the onset of the next trial, except in the case of 25% of filler trials, which included a bonus feedback page after clicking (see detailed explanation below). Figure 7.1 shows the complete flow of one example trial from the experiment.

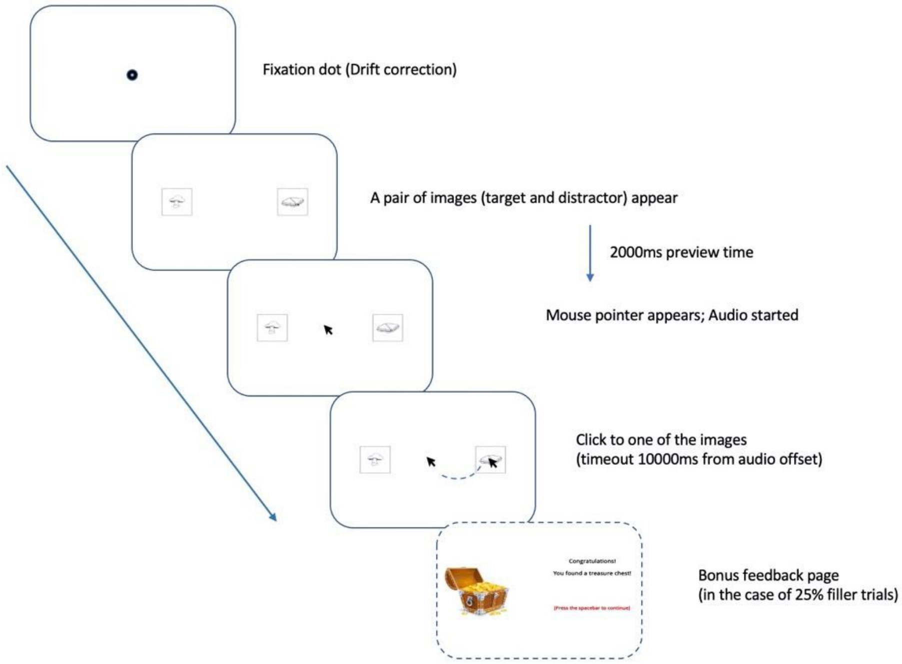


Figure 7.1 Procedure of a single trial from the experiment

Three aspects of the experiment were designed to ensure that participants remained motivated to discover the “true” location of the treasure throughout. The first is that participants were informed that their score would be counted based on how much treasure they managed to find, and that good scores would be entered on a high-score table which was shown at the end of the experiment. Second, participants were told that there were a certain number of bonus trials, from which they could earn more points than from other trials. To this end, 25% of the filler trials were immediately followed by a screen saying that a bonus treasure chest had been found, regardless of which picture had been clicked. To prevent learning, immediate feedback, such as this bonus message, was only provided in a small number of the filler trials, not in any critical trial, and only to keep participants motivated. Third, if no click was detected within 5 seconds of audio onset, a time-out message would be automatically shown on the screen to encourage participants to respond more quickly.

Eye movements were recorded during all experimental trials, together with the final object the participants clicked on. Only the data from critical trials were analysed.

Once the treasure-hunting part of the experiment had finished, participants filled out a paper version of the AQ test (Baron-Cohen et al., 2001). Finally, participants were fully debriefed about the aims of the experiment.

7.3 Results

Data and analysis scripts are available at https://osf.io/jwhfr/. Prior to analysis, practice and filler trials were removed, as well as the trials in which participants failed to click on either object within 5000ms of audio onset (0.5% of the critical data).

7.3.1 Autism Quotient

Across 62 participants, AQ ranged from 2 to 38, with a mean of 15.4 (SD 7.26) and a median of 15.0. For illustrative purposes only, participants were split at the median into two equal-sized groups, respectively referred to as the Low- and High-AQ groups. Characteristics of the participants in each group are shown in Table 7.1, and these groups will be used in tables and figures to illustrate effects.

Table 7.1 Characteristics of High-AQ group and Low-AQ group

	Mean AQ (SD)	Mean Age (SD)	Gender
Low-AQ Group	9.74 (3.04)	23.00 (0.5)	24F/7M
High-AQ Group	21.16 (5.53)	22.10 (0.5)	20F/11M

7.3.2 Final Object Clicks

Final object-click results were determined by two factors: whether the x-coordinate of the mouse when clicking the object is greater than 0, and the position of the target in that trial (such that if x>0 and the target is on the right, the target has been clicked). The percentages of trials in which participants clicked on targets and distractors, for the whole study and for the High- and Low-AQ participant groups, are given in Table 7.2.

Table 7.2 Percentage of mouse clicks recorded on each object (target/distractor) by manner of delivery (fluent/disfluent), for the whole study

	Whole study	Low-AQ	High-AQ

Fluent	Target	66.4	67.0	65.9
	Distractor	33.6	33.0	34.1
Disfluent	Target	50.2	50.3	50.2
	Distractor	49.8	49.7	49.8

The binary outcome of clicking on the target versus distractor was modelled in a mixed- effects logistic regression with one within-participants predictor of manner of delivery (fluent or disfluent) and one between-participant predictor of (raw, centred) AQ score, including random intercepts and a slope for manner of delivery by participant, as well as random intercepts by target image. Participants were 2.71 times more likely to click on the target following a fluent utterance compared to a disfluent utterance ($\beta = 0.9966$, SE = 0.3052, $p = 0.0011$). There were no effects of AQ ($\beta = -0.0226$, SE = 0.0256, $p = 0.3784$), or interactions between fluency and AQ ($\beta = 0.0121$, SE = 0.0421, $p = 0.7743$).

7.3.3 Eye Movement

Measuring the frequencies over time with which participants fixate each of the images presented, using the eye tracking record, allows us to investigate how biases are updated in the ongoing interpretation of what is said (see Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eber- hard, & Sedivy, 1995 for more details). In the present experiment, fixation data were averaged into bins of 20ms (10 samples). The proportions of time participants spent fixating objects were coded according to region of interest (target/distractor) for each bin. Figure 7.2 (top) shows the time course of all participants' fixations to target objects and distractors over 2000ms (100 bins) starting at target onset, for fluent and disfluent conditions respectively. For illustrative purposes, Figure 7.2 (bottom-left) and Figure 7.2 (bottom-right) show the equivalent patterns for the Low- and High-AQ groups of participants.

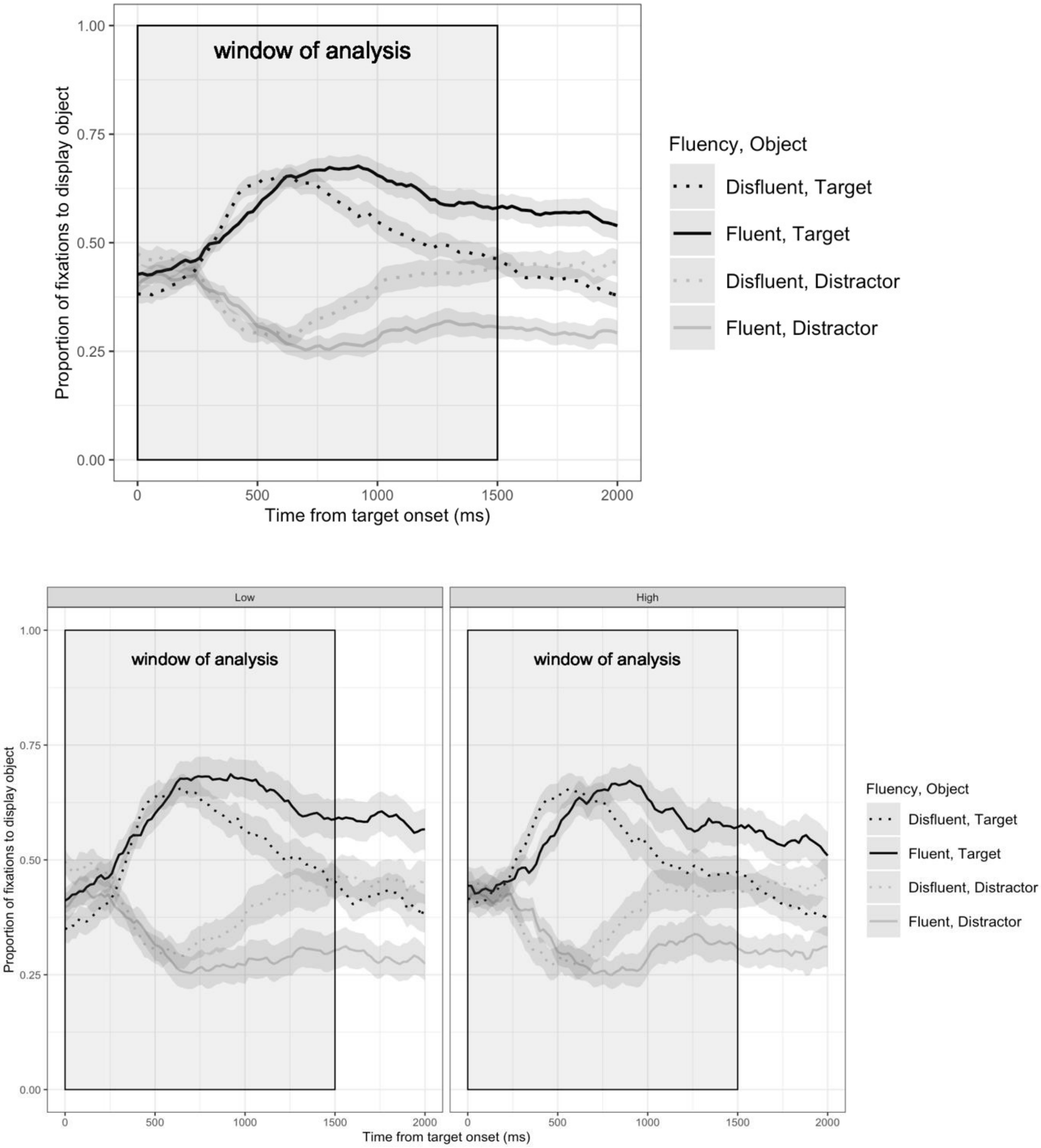


Figure 7.2 Mean proportions of fixations to target object and distractor over 2000ms from target onset, for fluent and disfluent conditions for all participants (top), Low- (bottom-left) and High- AQ (bottom-right) groups, calculated from summed fixations for each 20ms time bin from target onset to 2000ms post-onset. The shaded rectangular area represents the analysis window from target onset to 1500ms later. The shaded areas represent ±1 standard error of the mean by items.

We analysed eye movement data over a time window which began at target onset and ended 1500ms later. This window was chosen based on previous research suggesting that the resolution of a referent is normally established around 400-800ms after the object being mentioned (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Hanna, Tanenhaus, & Trueswell, 2003), with additional time allowed to capture any potential differences in later processing. For the purposes of analysis, we calculated a binomial "target advantage" for each bin: Bins in which the majority of time was spent fixating the target were coded as 1; bins in which the majority of time was spent fixating the distractor were coded as 0. Bins in which neither target or distractor were fixated (9,354) or in which both images were fixated equally (21) were discarded (10% of the data). Figure 7.3 shows target advantage, split by High- and Low-AQ groups, for reference.

We constructed a Bayesian generalised linear mixed model, with a logit link function, to analyse our data. We fit a model including orthogonal 1st-3rd order polynomial representations of time, allowing us to properly capture the non-linear nature of the target advantage shown in Figure 7.3. As well as the time variables, the model included a within- participants and -items effect of disfluency; and a between-participants, but within-items, measure of (scaled) AQ, as well as interactions between disfluency, AQ, and each of the polynomial orders of time. Random effects for intercepts and slopes (but not interactions) were included. The model was fit using the brms package in R (Bürkner, 2017), version 2.16.1, using default priors, with four chains of 4000 iterations each. The Rhat parameter for each parameter in the model was equal to 1, indicating successful model convergence; visual inspection of posterior predictive checks suggested a good fit to the data.

Details of the converged model are given in Table 7.3. For each effect we report a mean effect estimate, the upper and lower bounds of a 95% credible interval, and the probability that a given effect is different from zero in the direction of its sign. There is evidence for an effect of disfluency (the probability that participants fixate the target less overall when the utterance is disfluent is .986) and there are notable interactions of disfluency with the polynomials of time (effectively, indicating that the curve over time of the target advantage has a different shape when utterances are disfluent). Critical to the present paper are the three-way interactions between AQ, Fluency, and each of the polynomial orders of time. There is strong evidence for each interaction (all *p*s ≥ 0.992), showing that the target advantage curve changes with AQ, such that participants with higher AQ scores tend to show earlier biases towards the target. Figure 7.3 shows posterior predictions from the Bayesian model for the Low- and High-AQ groups, to illustrate this effect (but it is important to note that the effect is continuous, across the range of reported AQs).

Table 7.3 Model estimates for eye-tracking data, showing mean effect estimates in logits, upper and lower bounds of 95% credible intervals, and probabilities that effects differ from zero in the direction of their signs (based on posterior sampling).

effect	b	L95%	H95%	P (b > / < 0)
(Intercept)	0.76	0.51	1	—
AQ	-0.12	-0.34	0.1	0.863
Disfluency	-0.33	-0.64	-0.03	0.986
Time	2.53	1.21	3.85	1
$Time^2$	-2.13	-3.09	-1.23	1
$Time^3$	-0.54	-1.23	0.13	0.944
AQ:Disfluency	0.02	-0.23	0.27	0.552
AQ:Time	-0.46	-1.56	0.61	0.802
AQ:$Time^2$	0.08	-0.55	0.71	0.59
AQ:$Time^3$	0.12	-0.41	0.67	0.671
Disfluency:Time	-2.36	-2.63	-2.09	1
Disfluency:$Time^2$	-0.42	-0.7	-0.15	0.999

Disfluency:Time^3	1.36	1.09	1.63	1
AQ:Disfluency:Time	-0.61	-0.88	-0.33	1
AQ:Disfluency:Time^2	0.34	0.07	0.62	0.995
AQ:Disfluency:Time^3	0.33	0.07	0.6	0.992

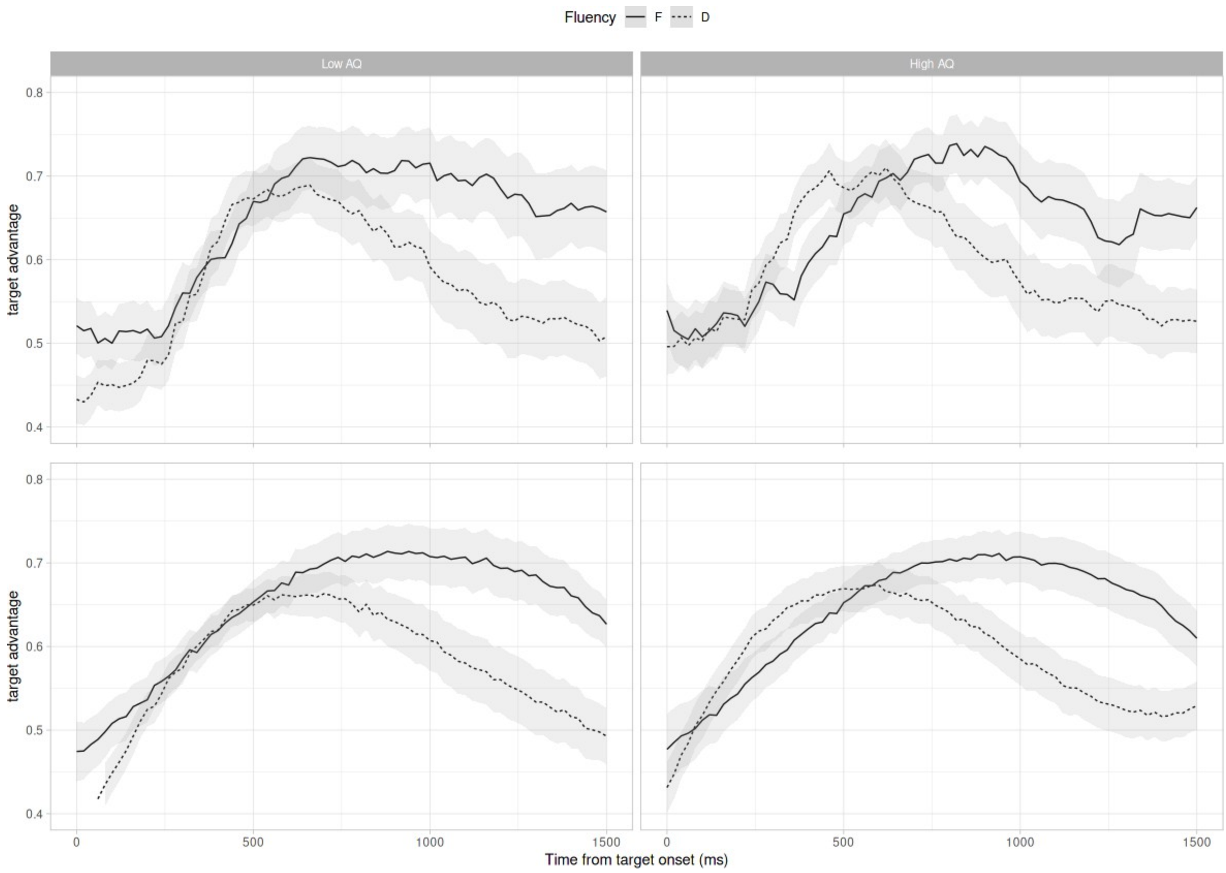


Figure 7.3 Recorded (top) and model-predicted (bottom) target advantage, for Low-AQ (left) and High-AQ (right) groups of participants. Model predictions are derived from 101 posterior samples from the converged model.

7.4 Discussion

The results reported here confirm that listeners' pragmatic judgements are influenced by manner of speech, and that disfluent instructions are associated with deception. We took the greater frequency of clicks on the distractor object after hearing an instruction such as "The treasure is behind, thee, uh, hat" as evidence that participants were less inclined to believe the location of the treasure named by the speaker when the utterance was disfluent. Moreover, eye movements made during the early time-course of comprehension showed that these judgements were made quickly. These findings replicate the results of Loy et al. (2017).

Note that in these studies, participants were explicitly told that they might be deceived, and the utterances to which they responded were recorded. Whether there are general conclusions to be drawn about deception is unclear; here we are concerned with the processing of language that is 'known' not to have a literal interpretation. We can however draw some ecological comfort from a study by Loy et al. (2018), in which participants' judgements were affected in a similar way by disfluency in a two-player version of the current deception game, in which all utterances were spontaneous.

Critically, the present study shows that the general effects of disfluency described above vary by AQ. When hearing disfluent instructions, participants with higher AQ scores, and therefore higher numbers of traits typically associated with autism, are more likely to make early fixations on the target. The evidence for these effects is manifest as a series of interactions with a continuous AQ score predictor, and therefore not subject to criticisms typically applied to median splits (e.g., McClelland, Lynch, Irwin, Spiller, & Fitzsimons, 2015). However, the differences emerge as differences of processing but not of comprehension: Our data do not suggest that participants' final interpretations vary with AQ, but they do suggest that there is variation in the mental processes which underlie those interpretations.

A potential account of these findings is that as the number of autistic-like traits associated with higher AQ scores increases, participants have more difficulty in utilizing contextual information. This may lead to a weaker association between disfluency and deception, attenuating the early tendency to fixate the distractor following a disfluent utterance. A variant of this account is that it may be more difficult to override more basic effects of filler disfluencies, such as their tendency to focus listeners' attention on the upcoming message (Collard, Corley, MacGregor, & Donaldson, 2008; Fox Tree, 2002). According to the latter view, participants with higher AQs may show an initial tendency to fixate the target quickly following a disfluency, before this is overridden by slower pragmatic reasoning.

It is important to stress again that these findings are only relevant to people diagnosed with autism to the extent that AQ measures can be considered a proxy for diagnosis. The critical aim of the paper is not to identify a processing difference in the clinical population, but to point to a likely candidate which merits further investigation. Moreover, because of the continuous nature of the AQ, some statistical issues which can arise from group comparisons (especially when “controlling” for other varying factors: Miller & Chapman, 2001) are avoided. We can be confident (to the usual limits of statistical inference) that the effect is real: in the subclinical population, comprehension processes may result in the same outcomes, even though the processes by which those outcomes are achieved may vary.

In line with the study by McKenna et al. (2015), the implication of the current study is that by merely focusing on the outcomes of cognitive processing, it is possible to miss the real differences in detailed processes that vary across the population. Previous studies have used tasks in which participants are asked to provide context-appropriate explanations for non-literal speech, such as Happé’s Strange Stories Test (Happé, 1994), in exploring disfluency comprehension in ASD children. However, many of these studies represent cases in which ASD- and non-ASD individuals achieved similar results, despite the ASD participants being expected to show less developed ability (Hala et al., 2007; Happé, 1995; Jolliffe & Baron- Cohen, 2000). Such results may not be conclusive without exploring the detailed cognitive processes underlying any parity in behavioural outcomes.

The present study suggests that there are differences in the processing of disfluency among individuals varying in numbers of self-identified autistic traits. This may, by extension, have implications for the comprehension processes of people with ASD diagnoses.

7.5 Conclusions

This chapter presents an eye-tracking experiment investigating whether the pragmatic reasoning process cued by disfluency vary across individuals. The current experiment was conducted with individuals whose pragmatic abilities were posited to vary, suggested by Autism Quotient (AQ). The eye-tracking data revealed an interaction between the speaker’s manner of delivery and AQ, showing that when hearing disfluent utterances, a tendency to make early fixations to the named object increased with individuals’ AQ scores. These findings suggest that the reasoning process engaged in comprehending disfluencies is related to listeners’ pragmatic abilities, providing further evidence that disfluency processing involves a social reasoning process, rather than a heuristic association. Critically, the current study highlighted the understanding of process in language comprehension, as even where utterances are equivalently understood, the process by which interpretations are achieved varies across individuals.

Chapter 8
General Discussion

This thesis described a series of experiments that investigated the effects of the disfluency *uh* on language comprehension. A combined effect of disfluency and social contexts on listeners' pragmatic comprehension of ambiguous statements was assessed using mouse-tracking experiments (Experiments 1-4). A systematic variation of individuals' pragmatic reasoning in disfluent speech comprehension was assessed with an eye-tracking experiment (Experiment 5). In this chapter, I summarise the main findings from the experiments, propose a possible account to explain the results, and suggest potential directions for future research.

8.1 Summary of the main findings

This summary includes two sections. The first section summarises the results of four mouse-tracking experiments (Experiments 1-4), which investigated the effects of the disfluency *uh* on listeners' interpretations of the ambiguous scalar word *some*. The second section summarises the results of Experiment 5, which investigated individual differences in disfluent speech comprehension.

8.1.1 Summary: Experiments 1-4

A series of mouse-tracking experiments were conducted to test how the disfluency *uh* influences listeners' interpretations of ambiguous utterances with the scalar word *some*. The mouse-tracking results, including participants' final interpretations of *some* and their real-time mouse-movements, showed a combined pragmatic effect of *uh* and of social contexts. Listeners were able to rapidly integrate the speaker's communication goals and the given social contexts as utterances were processed to reach an ultimate interpretation of *some*. Specifically, in a context where larger values were socially undesirable (e.g. "My vehicle has failed *some* inspections"), listeners were biased towards a literal interpretation of *some* as "all" following disfluent utterances (Experiments 1-2). This result was consistent with that in Loy, Rohde and Corley (2019). However, when varying the context to one where smaller values are socially undesirable (e.g., "I've got *some* 'A's for my psychology modules"), disfluency showed an opposite effect, reducing the values participant associated with *some* to "one" or even "zero" (Experiments 3-4).

Critically, mouse-trajectory analyses showed that listeners were quick to commit to interpreting *some* to "one" (Experiment 3) and slow to commit to a "zero" interpretation (Experiment 4). This result suggested that when encountering disfluencies, extending the meaning of some to "one" was relatively easy, but overwriting the meaning to "zero" (i.e., deciding the speaker is lying) was more demanding.

8.1.2 Summary: Experiment 5

An eye-tracking experiment was conducted with individuals whose pragmatic abilities were assumed to vary, suggested by Autism Quotient (AQ). This experiment was to test whether the pragmatic reasoning processes cued by disfluencies vary across individuals. Listeners needed to click on one of two objects in order to find some hidden treasure, based on a speaker's fluent or disfluent utterance naming the location of the treasure (e.g., "The treasure is behind thee, uh, hat"). Both the final clicks and participants' moment-by-moment eye movements were recorded. The results showed fewer clicks on the mentioned target object following disfluent utterances, in keeping with prior work (Loy, Rohde, & Corley, 2017).

Importantly, the eye-tracking data revealed an interaction between manner of delivery and AQ. When hearing disfluent utterances, an early bias towards a literal interpretation correlates positively with AQ. This suggests that, even with similar final click outcomes whereby disfluency leads to fewer target interpretations for individuals with various AQ, the process by which the interpretation is achieved varies across individuals.

8.2 Social reasoning: an account of disfluency processing

The current experiments support an account of disfluency processing whereby listeners rapidly integrate pragmatic cues from disfluency into a social reasoning process in a given context, a process varying with listeners' pragmatic abilities.

8.2.1 Two stages of reasoning

Results from both mouse-tracking and eye-tracking experiments can potentially be encompassed by a two-stage social reasoning account of disfluency processing. This account assumes two stages of comprehending, (1) in the first stage, listeners reach an initial interpretation following disfluency and (2) in the second stage of pragmatic reasoning, listener override the initial interpretation if the social context allows.

For example, in the mouse-tracking experiments, this two-stage social reasoning was manifested in a way that listeners applied both strategies of stretching and overwriting when searching for an interpretation of the ambiguous word *some*. The social context of job interviews set up an expectation that the speaker might seek to misleading the listener. Therefore, when lower values were the undesirable interpretations of *some*, listeners encountering disfluencies attempted to extend the boundary of the meaning of *some* as low as possible. The stages of reasoning were shown to comprise two stages in that in the first stage, listeners quickly arrive at an initial interpretation when the lowest number available still fits within the semantic boundary of "some" (e.g., “one”). However, in the second stage, when the lowest number available is outside this boundary (e.g., “zero”), listeners hesitate and engage in deeper social reasoning before potentially overriding the interpretation to zero.

This two-stage reasoning account also stems from one potential explanation for our eye-tracking experiments. Listeners with higher AQ showed earlier fixations to the named object compared to those with lower AQ. This result can actually be due to listeners with higher AQ experienced more difficulty in overriding a basic effect of disfluency which cues an initial attentional focus on the upcoming messages (e.g., Collard, Corley, & MacGregor, 2008). The two stages of disfluency comprehension in this experiment could involve this initial attention to the named object and this initial response is adjusted by a more time-consuming pragmatic reasoning process.

These proposed stages should be differentiated from previous staged models of implicature processing, such as the two-stage account of scalar implicature processing in Tomlinson Jr. et al. (2013). In our current studies, the lower bound meaning of *some* is undermined. Pushing implicatures to their limits and beyond allows us to answer a different question from that most commonly associated with scalar implicature (e.g., whether individuals first access the literal meaning of *some* "some and possibly all" and enrich it to form the scalar implicature "some but not all", or whether the implicated meaning is understood by default). Here, we emphasise the stages of how listeners incorporate the pragmatic information from disfluency *uh* to reason and adjust in the given social context to reach a final interpretation of *some*, which may happen **after** any "some but not all" implicature has been calculated and the meaning is additionally constrained to "some and possibly very few".

In the following sections, I detail the pragmatic effects of disfluency *uh*, as well as the role of social contexts, found in the current studies.

8.2.2 Pragmatic effects of disfluency *uh*

Facilitating comprehension, not hindering

With the research focus on filler *uh*, the studies in this thesis emphasised the pragmatic effects of disfluency in language comprehension. Disfluencies are able to facilitate, instead of hindering, listeners' processing of spontaneous speech.

The current studies showed the pragmatic effects of disfluencies in listeners' language comprehension in real time. In the mouse-tracking experiments, listeners showed a choice pattern consistently relating disfluency to the socially undesirable interpretations of *some*. The eye-tracking experiment showed a consistent pattern where listeners associated disfluent utterances with lying. These reported results further assured that instead of being mere interruptions of speech flow, disfluencies are able to facilitate listeners' language comprehension to reach an ultimate interpretation via acting as pragmatic cues. Critically, the current results showed an early effect of disfluency in processing. These results indicate that listeners are able to take social contexts and disfluency into processing quickly and they draw pragmatic inferences at the early stage of comprehension.

Our findings are consistent with previous findings that disfluencies are utilised as pragmatic cues by listeners to judge the speaker's state (e.g., Brennan & Williams, 1995; Zuckerman, Koestner, & Driver, 1981), or to heighten attention of the following utterances (e.g., Collard et al., 2008). The early effect of disfluency in processing found in the current studies is also consistent with evidence from previous studies that listener make rapid inferences about the speaker or the utterances (e.g., Grodner & Sedivy, 2011; Loy, Rohde, & Corley, 2017; Loy, Rohde, & Corley, 2019).

Even though disfluencies "facilitate" comprehension by helping listeners reach an eventual interpretation via pragmatic cues, our results indicate that disfluencies actually "betray" speakers by revealing their unintended intentions, such as potential deception. A possible perspective to interpret this finding is that disfluencies serve as a social contract, signalling when a speaker is engaging in exaggeration or presenting half-truths. Disfluencies can act as hedges, indicating a speaker's uncertainty or a strategic pause to manage the listener's perception (Corley & Stewart, 2008; Fraser, 2010; Gósy, 2001).

Perspective-taking, not heuristic

Importantly, the current results suggest an account which supports a perspective-taking account of disfluency processing instead of a heuristic account. The perspective-taking account expects that the pragmatic cues involved in disfluencies are context-specific and speaker-specific, as this account assumes the listener to interpret utterances via a speaker-modelling process. However, the heuristic account assumes an automatic processing of disfluency to certain pragmatic cues.

Our findings support the perspective-taking account on two fronts. First, when varying the given social contexts in mouse-tracking experiments, listeners interpreted the same scalar word *some* following disfluencies in different directions. In a social context where larger values of *some* are undesirable, listeners associate disfluency to larger-value interpretations of *some*. However, when varying the social context to one where smaller values are undesirable, disfluency showed an opposite effect that it reduced the values listeners associated with *some*.

If combining the discussion from the previous section, despite being different interpretations of *some*, these interpretations were all the socially undesirable ones in the given contexts. These results indicate that disfluency does not provide pragmatic cues to a certain interpretation of ambiguous words, but rather cues a social reasoning process where listeners take the speaker's conversation goals into consideration. For example, in the given social context of job interviews, the speaker's goal is to leave good impression and listeners consider possible reasons of the speaker's disfluencies in processing to reach an eventual interpretation. This finding is also consistent with broader evidence that comprehenders' interpretations of scalar word *some* are context-dependent (e.g., Bonnefon, Feeney, & Villejoubert, 2009; Breheny, Katsos, & Williams, 2006; Degen & Tanenhaus, 2014).

The second reason to support a perspective-taking account is that the disfluency comprehension process varies with individuals' social cognitive profile. Our eye-tracking results showed a general effect of disfluency influencing listeners' pragmatic judgements and disfluent utterances were associated with deception. The pragmatic effect of disfluency as cues for deception aligns with previous findings (e.g., King, Loy, & Corley, 2017; Loy, Rohde, & Corley, 2018; Zuckerman, DePaulo, & Rosenthal, 1981). Importantly, this effect varies by AQ: listeners with higher AQ, therefore higher numbers of traits typically associated with autism, are more quickly in fixating on the named object. This result indicates that associating disfluency to lying is not an automatic process cued by the occurrences of disfluency, but a reasoning process that varies by individuals' pragmatic abilities.

Contextual, not stand-alone

Our findings also re-emphasise the crucial role of social contexts in resolving pragmatic ambiguity. Pragmatic ambiguity is unique because the words involved do not always have an explicitly distinct meaning. Moreover, utterances with this kind of ambiguity do not require listeners to reach an agreed-upon meaning for successful comprehension, as listeners are usually sensitive to the vagueness if a speaker deliberately delivers an ambiguous utterance. These features distinguish pragmatic ambiguity from other types of ambiguity (lexical or syntactic) and indicate that this kind of ambiguity is context-sensitive.

Building on prior work (e.g., Bonnefon, Dahl, & Holtgraves, 2015; Loy, Rohde, & Corley, 2017; Loy, Rohde, & Corley, 2019), the current studies explored various social contexts, attempting to associate disfluencies with social situations such as face-threat, face-boost, or deception. Through this broad exploration of different social contexts, this thesis highlights the significant role of social context in resolving pragmatic ambiguity across various experiment paradigms. Listeners use context to integrate and interpret ambiguous utterances rapidly, underscoring the dynamic interplay between linguistic cues and contextual information in real-time language comprehension.

8.3 Implications for future research

The experimental work involved in this thesis highlights the importance of considering the detailed process of comprehension. Previous studies have shown that for a task where participants were asked to provide explanations for non-literal speech that fit the context, individuals with Autistic Spectrum Disorder (ASD) and those without showed similar performance (Hala, Pexman, & Glenwright, 2007; Happé, 1995; Jolliffe & Baron-Cohen, 2000). However, these two groups of individuals might achieve similar results via different processes. Mouse-tracking data in Chapter 6 showed there were similarly increasing numbers of clicks for one- and zero-qualification images following disfluent utterances. However, mouse-trajectory analysis revealed that listeners reached these two choices via different processes. Same for the eye-tracking movements in Chapter 7, where listeners with various pragmatic abilities showed differences in the comprehension process while the comprehension outcomes did not differ.

Future research should also be aware of the necessity to looking into the detailed process underlying comprehension rather than just the outcomes. For example, given the account of social reasoning concluded in the mouse-tracking experiment, it is encouraged to explore new questions about individual difference of this social reasoning process in interpreting ambiguous word *some*, or the processing differences when varying social contexts which involve different conversational goals.

The current findings in Experiment 1-4 showed that a basic social reasoning occurs early which forms a fast initial stage of reasoning. However, when social reasoning conflicts with lexical meaning (i.e., overwriting the meaning of *some*), it takes longer and therefore form the second stage of pragmatic processing in a given social context. Besides the processing time, there may exist more detailed differences between these two reasoning processes. For example, when the meaning of *some* is stretched (Experiment 3), the reasoning process might require listeners to resolve a

morphosyntactic conflict between the word *some* and the plural form of the noun (e.g., languages). Although time consuming, this conflict might still be faster to resolve compared to more severe conflicts with meanings, such as a **semantic conflict** in Experiment 4. Where the intended meaning of *some* directly conflicts with its semantic meaning (Experiment 4), it requires a deeper reasoning process where listeners overwrite the meaning of *some* and choose an interpretation that is at odds with the semantic meaning of *some*. To further distinguish these two stages, we suggest replicating the current mouse-tracking studies in a language that does not have a plural form of the noun (e.g., Chinese or Korean), so that morphosyntactic conflict does not occur and a more explicit difference could be observed between these two stages.

The present findings are obviously limited to the filler type of disfluency *uh*, the scalar word *some*, and the individual differences suggested by AQ. Questions remain whether the currently proposed two-stage account of disfluency processing can also be observed using other types of disfluencies (e.g., repetitions or lexical fillers “you know”), other scalar words (e.g., a few), or other indicators of individual differences (e.g., gender). For example, can listeners interpret *a few* to “one” if given the same context as in the current studies? Will we observe different processing, such as different patterns in clicking results or trajectories, compared to the current studies using *some*? And can listeners interpret *a few*, a scalar word which by definition should refer to a small number, to mean a larger number in a given social context? Or other questions like: using the current mouse-tracking data, will we observe any gender differences in processing? And will the gender of the speaker also influence the processing of listeners in the given social context of job interview? These are all research questions that can be further explored to validate the two-stage account.

Methodologically, the mouse-tracking experiments in this thesis show the possibility of conducting web-based mouse-tracking task to explore listeners’ real-time language processing. The benefit of this methodology is easy implementation and fast participant recruitment. Moreover, this thesis used a four-item mouse-tracking paradigm, different from a typical two-item mouse-tracking paradigm (Spivey, Grosjean, & Knoblich, 2005). This demonstrates the potential for future research to utilise more diverse visual representations to explore time-course movements. However, the analysis of mouse-tracking data is still underexplored and may need more attention.

8.4 Conclusions

The series of experiments reported in this thesis focused on investigating the effects of the disfluency *uh* in socially-situated real-time language comprehension. Combining four mouse-tracking experiments and one eye-tracking experiment, the current studies support an account in which disfluency cues a rapid social reasoning process where listeners integrate the pragmatic cues from disfluency fast in a given social context; a process which varies with individuals' pragmatic abilities. Specifically, this process includes two stages of reasoning that listeners reach an initial interpretation compatible with the broad semantics of the utterance, but can then override this if context and disfluency allow.

Appendix A

Li, Rohde, & Corley (2020)

Journal of Autism and Developmental Disorders (2022) 52:4921–4930
https://doi.org/10.1007/s10803-021-05347-4

ORIGINAL PAPER

Veritable Untruths: Autistic Traits and the Processing of Deception

Wei Li[1] · **Hannah Rohde**[2] · **Martin Corley**[1]

Accepted: 29 October 2021 / Published online: 18 November 2021

Abstract

How do we decide whether a statement is literally true? Here, we contrast participants' eventual evaluations of a speaker's meaning with the real-time processes of comprehension. We record participants' eye movements as they respond to potentially misleading instructions to click on one of two objects which might be concealing treasure (*the treasure is behind thee, uh, hat*). Participants are less likely to click on the named object when the instructions are disfluent. However, when hearing disfluent utterances, a tendency to fixate the named object early increases with participants' autism quotient scores. This suggests that, even where utterances are equivalently understood, the processes by which interpretations are achieved vary across individuals.

Keywords Disfluency · Autism spectrum quotient · Eye-tracking

Introduction

Even though autism spectrum disorder (ASD) is generally associated with a deficit in social and figurative aspects of communication, many previous studies investigating the abilities of participants with ASD to understand non-literal language have shown that these participants obtain similar comprehension scores to their neurotypical peers (e.g., Happé, 1994, 1995; Hala et al., 2007; Jolliffe & Baron-Cohen, 2000; Loukusa & Molianen, 2009). This raises the question of whether individuals with ASD and those without achieve similar outcomes by employing different processes. Here, we develop the theme of processing differences, with a focus on how the manner of a spoken message is delivered, in particular, disfluent fillers such as "um" and "uh", influences individuals' comprehension.

Fillers are often regarded as socially oriented, influencing discourse dynamics by controlling the flow of conversation (e.g., Clark & Fox Tree, 2002), signalling hedges in conversation (Smith & Clark, 1993), or being taken as cues to deception (Loy et al., 2016; Vrij & Semin, 1996; Zuckerman et al., 1981). Previous studies have shown that high-functioning individuals with ASD produce fewer fillers than do the typical population (Lake et al., 2011). Children with autism produce fewer fillers compared to their typically developing peers (Gorman et al., 2016; Heeman et al., 2010; Irvine et al., 2016). This reduced use of fillers, considered as a conversational cue, is consistent with a characterization of speakers with ASD as being less able to take their listeners' perspectives (Colle et al., 2008; Paul et al., 2009). A decreased attention to social cues may yield difficulty in understanding the implicit social meanings associated with fillers (Irvine et al., 2016).

However, this difficulty may not manifest in comprehension outcomes. Participants who struggle with social cues may nonetheless find ways to interpret fillers in an equivalent way to their peers. By focusing on the outcome of comprehension, usually operationalised as whether the eventual interpretation of an utterance aligns to a norm, it is possible that differences in the processes which underlie these outcomes may be overlooked (McKenna et al., 2015). McKenna et al. (2015) demonstrated a difference in the processing of a type of figurative language, metonymy, in which an attribute of a concept instantiates that concept (such as the use of *Dickens* when referring to the books written by Charles Dickens rather than to the writer himself). In their study, participants without diagnoses completed an autism spectrum quotient questionnaire (AQ: Baron-Cohen et al., 2001)

✉ Martin Corley
Martin.Corley@ed.ac.uk

[1] Psychology, PPLS, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

[2] Linguistics and English Language, PPLS, University of Edinburgh, Dugald Stewart Building, Edinburgh EH8 9AD, UK

and read sentences containing target nouns (e.g., *Vietnam, Finland*), in contexts supporting literal or figurative, metonymic, interpretations (e.g., *During my trip, I hitchhiked around Vietnam/Finland; A lot of Americans protested during Vietnam/Finland*). Their results showed an "unfamiliar metonym disadvantage" where participants were slower to read a novel metonym (e.g., *Finland*) in its figurative sense, compared to a metonym which has its figurative meaning established in daily use (e.g., *Vietnam* referring to the conflict during the 1950s–1970s). Of relevance to the present study, they found that participants with higher scores on the autism quotient measure were relatively slowed down by these novel metonymies, while those with lower numbers of self-reported autistic traits were not. Note that all participants scored 100% on comprehension scores: differences related to AQ could not be found in any measure of interpretation, but were instead only evident when the time course of the processing of the stimuli was investigated.

Although understanding figurative meanings in written language, such as that investigated by McKenna et al. (2015), can be difficult, understanding non-literal meaning in speech adds additional complexity, because artefacts such as fillers can modify literal meaning. Loy et al. (2017) investigated whether interpretations of a sentence (truthful or deceptive) made by listeners in the general population could be influenced by disfluency. Their experiments used a lie-detection game setting, where participants were presented with two images (named image/distractor image) in every trial and were asked to click on the image that they thought concealed some treasure. For each trial, participants heard a pre-recorded fluent or disfluent utterance describing the location of the treasure. Participants' image clicks revealed a bias to interpret disfluent speech as a cue to speaker deception about the location of the treasure. Importantly, eye movements were also measured, and disfluent speech resulted in more frequent early eye movements towards the distractor item, showing the influence of an early process associating disfluency with deception.

Loy et al. (2017) study provides two essential points of reference for the current study. First, the method allows us to directly observe the time-course of disfluency comprehension, providing a way to investigate how the manner in which something is said might affect its ongoing interpretation. Second, along with other situations which depend on understanding the interlocutor's intentionality, the recognition of deception has been shown to be difficult for individuals with ASD: for example, it is harder for children with autism to lie to others and to detect when they are being lied to (Ranick et al., 2013). By using a lie detection paradigm in which disfluency can serve as a potential pragmatic cue, we can examine in detail any differences in the processes of comprehension that might arise between individuals. Specifically, we can measure both participants' eventual interpretations—truth or deception?—and the real-time processes that precede that outcome.

Our experiment replicates experiment 2 from Loy et al. (2017), with a set of participants from the general population who vary in their autism spectrum quotients (AQ). AQ is not a diagnostic test of ASD, but it is used to assess broader autism phenotype (BAP) metrics for understanding subclinical ASD traits within individuals in the general population who are presumed to have typical developmental profiles. The test comprises 50 questions focusing on five BAPs: social skills, attention switching, attention to detail, imagination, and communication (Baron-Cohen et al., 2001). AQ scores from low to high quantify where a given participant is situated on a putative continuum from neurotypical to autism diagnosis (Baron-Cohen et al., 2001). While our findings do not allow us to draw direct conclusions about comprehension processes in ASD, the study emphasizes the utility of real-time measures in comparing processing differences among populations that vary on an AQ spectrum.

Our experiment measures the influence of manner of delivery on pragmatic comprehension in two ways. First, we measure the *outcome* of the comprehension process by recording which object people click on. This gives us a direct indication of whether a participant has interpreted a given utterance as truthful, in which case they are likely to click on the named (target) object, or deceptive, in which case they are likely to click on the other (distractor) image. If the results of Loy et al. (2017) are replicable, we expect to see an effect of the presence of disfluency, such that participants are less likely to click on the target following a disfluent utterance. We further assess the *process* of comprehension. By measuring participants' eye movements time-locked to the speaker's mention of the target object, we can test the time-course of comprehenders' responses to the speaker's manner of delivery. If the results replicate those of Loy et al., we should see a decrease of bias towards the named object in the disfluent condition. Of critical interest in the present study is whether the outcome or process measures interact with AQ. An interaction at the outcome level (item clicked) would suggest that the levels of broader autism phenotypes measured by the AQ affect the eventual understanding of what is said. Regardless of whether such an interaction is found or not, an interaction at the process level (eye movements) would suggest differences in the processes underlying non-literal interpretations, opening the door to research with formally-diagnosed groups of participants.

Methods

Participants

Ethical approval was given by the PPLS Research Ethics Committee with reference number 100-1920/1. Sixty-two self-reported native English speakers (Male: 18; Female: 44) took part in the experiment. Participants ranged from 18 to 35 years old, mainly comprising students at the University of Edinburgh recruited from a wide variety of disciplines, including Engineering, Chemistry, Mathematics, Law, and Psychology. Participants were paid £5 for participation. Participants had normal or corrected-to-normal vision, and all used their right hands to control the mouse used in the experiment. Participants provided written consent prior to the beginning of the experiment.

Material

Autism Spectrum Quotient

The autism spectrum quotient (AQ: Baron-Cohen et al., 2001) questionnaire comprises 50 items reflecting five different broader autism phenotypes either positively (e.g., "I am fascinated by numbers", "I am often the last to understand the point of a joke"), or negatively (e.g., "I find social situations easy"). Participants are asked to respond based on a 4-point Likert scale varying from "definitely agree" to "definitely disagree". Each response represents one point; and a participant's AQ score is calculated by totalling their responses (taking polarity into account), with a maximum score of 50. Studies to date suggest that individuals diagnosed with autism tend to score 32 or more, compared to individuals without autism who tend to average around 16. However, the AQ is not a diagnostic test, and autistic traits are present in the wider community (see Folstein & Rutter, 1977; Persico & Bourgeron, 2006 for details). Therefore, AQ scores should be interpreted very carefully. Participants in the current study were informed of this beforehand and were told that their AQ scores would not be provided to them.

Visual Stimuli

The visual world paradigm from Loy et al. (2017, experiment 2) was adapted for the current study. Visual stimuli included 130 line-drawings from Rossion and Pourtois (2004), which were grouped into pairs across 65 trials (5 practice trials; 20 critical; 40 fillers). For each pair of drawings in each trial, the drawing that the speaker named as concealing the hidden treasure is referred to as the *target* picture, and the other one the *distractor* picture. The two pictures were presented vertically centred on the screen, and horizontally centred at 25% (left-hand picture) and 75% (right-hand) of the screen width. The target picture's position on the left versus the right was counterbalanced across items.

Audio Stimuli

The audio files used in the current experiment were taken from Loy et al.'s Experiment 2. Participants were told, as a cover story, that the speaker had taken part in an earlier experiment, in which she had been told to try and mislead her partner about the treasure's location by sometimes lying.

In 20 critical trials, the recordings of the speaker describing the purported locations of the 'treasure' were categorized as either *fluent* (e.g., "The treasure is behind the [target]") or *disfluent* (e.g., "The treasure is behind thee, *uh*, [target]"). To ensure that participants were exposed to the same utterance (bar disfluency) across conditions, and the same disfluency across disfluent trials, complete fluent and disfluent sentences were first recorded, and then a prolonged article followed by a filler cut out from a disfluent utterance was spliced into each relevant fluent utterance to create the corresponding disfluent version. Disfluent utterances were therefore characterized by a prolonged article ("the" pronounced /ðiː/, or "thee") and the disfluent segment (i.e., "uh") before the target noun.

The assignment of fluent/disfluent conditions to items was counterbalanced across two lists. However, due to a coding error, one critical item ("lamp") in List 1, which should have been in the fluent condition, was mistakenly paired with a disfluent recording. Thus, the final numbers of items in experimental conditions for participants using audio files from List 1 were $n = 9$ (fluent items) and 11 (disfluent items).

To obscure the experimental manipulation, and reinforce the impression that participants were listening to natural recordings, forty additional filler trials referred to pairs of images that were not used in critical trials, and were designed such that half included disfluencies other than "uh", or discourse manipulations such as discourse markers (e.g., "*Okay*, the treasure is behind the [target]") and modals (e.g., "The treasure *could be* behind the [target]"), and the other half were fluent.

Procedure

The experiment was in two parts, a computer-based lie detection game and an Autism-Spectrum Quotient (AQ) measurement. The researcher began by introducing the process of the whole experiment and the lie detection game task. The eye-tracking experiment which followed was presented using OpenSesame version 3.2.7 (Mathôt et al., 2012), with movements of the right eye monitored using an Eyelink 1000 Tower Mount system. Participants were seated at a viewing distance of 80 cm from a 19″ CRT monitor. They read the instructions first which explained that their aim was to collect as much treasure as possible by clicking on the object concealing the treasure on each trial.

Following reading the brief instructions, a press of the spacebar started the calibration of the eye-tracker. Five practice trials followed the calibration; at the end of the practice session, participants pressed the space bar to commence the sixty experimental trials (20 critical trials and 40 fillers).

Each new trial began with a black fixation dot at the centre of the screen, and participants were told to carefully stare at the black dot when they were ready to start a trial, allowing a drift correction to be applied if necessary. As soon as the fixation dot disappeared, a pair of images—a target and a distractor—appeared on the screen. After a preview period of 2000 ms, the mouse pointer appeared at the centre of the screen, and the audio stimulus started. Participants were instructed to click on the object which they believed the treasure was behind as soon as they could. Once one object had been clicked on, the images disappeared, providing visual feedback that the click had been recorded. 1000 ms after the end of the audio, or once the participant had clicked on an object—whichever occurred later—a black dot appeared which indicated the onset of the next trial, except in the case of 25% of filler trials, which included a bonus feedback page after clicking (see detailed explanation below). Figure 1 shows the complete flow of one example trial from the experiment.

Three aspects of the experiment were designed to ensure that participants remained motivated to discover the "true" location of the treasure throughout. The first is that participants were informed that their score would be counted based on how much treasure they managed to find, and that good scores would be entered on a high-score table which was shown at the end of the experiment. Second, participants were told that there were a certain number of bonus trials, from which they could earn more points than from other trials. To this end, 25% of the filler trials were immediately followed by a screen saying that a bonus treasure chest had been found, regardless of which picture had been clicked. To prevent learning, immediate feedback, such as this bonus message, was only provided in a small number of the filler trials, not in any critical trial, and only to keep participants motivated. Third, if no click was detected within 5 s of audio onset, a time-out message would be automatically shown on the screen to encourage participants to respond more quickly.

Eye movements were recorded during all experimental trials, together with the final object the participants clicked on. Only the data from critical trials were analysed.

Once the treasure-hunting part of the experiment had finished, participants filled out a paper version of the AQ test (Baron-Cohen et al., 2001). Finally, participants were fully debriefed about the aims of the experiment.

Results

Data and analysis scripts are available at https://osf.io/jwhfr/. Prior to analysis, practice and filler trials were removed, as well as the trials in which participants failed to click on either object within 5000 ms of audio onset (0.5% of the critical data).

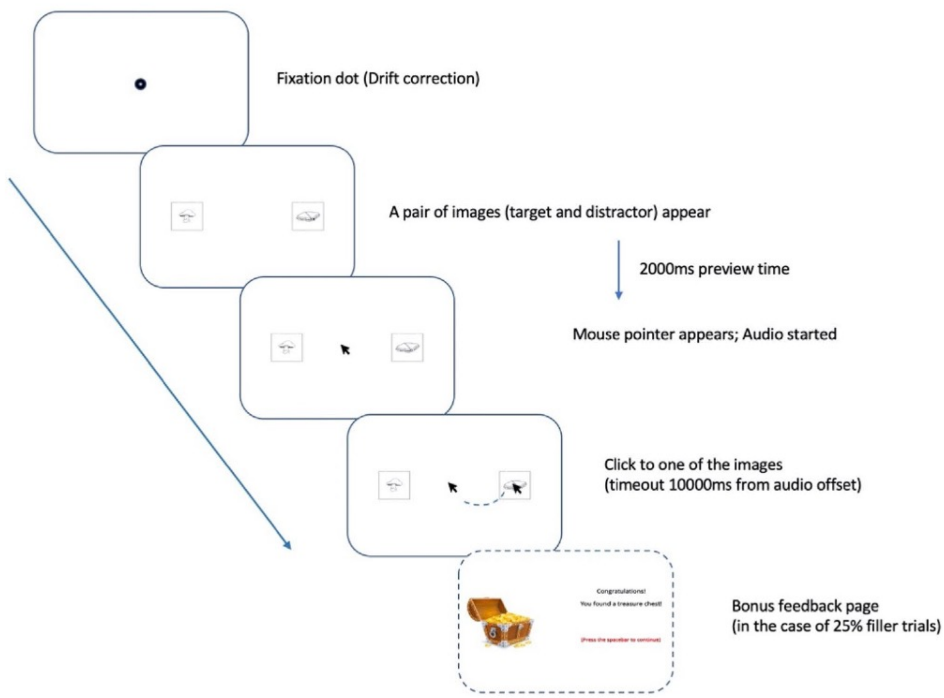


Fig. 1 Procedure of a single trial from the experiment

Table 1 Characteristics of High-AQ group and Low-AQ group

	Mean AQ (SD)	Mean age (SD)	Gender
Low-AQ group	9.74 (3.04)	23.00 (0.5)	24F/7M
High-AQ group	21.16 (5.53)	22.10 (0.5)	20F/11M

Table 2 Percentage of mouse clicks recorded on each object (target/distractor) by manner of delivery (fluent/disfluent), for the whole study

	Whole study	Low-AQ	High-AQ
Fluent			
Target	66.4	67.0	65.9
Distractor	33.6	33.0	34.1
Disfluent			
Target	50.2	50.3	50.2
Distractor	49.8	49.7	49.8

Autism Quotient

Across 62 participants, AQ ranged from 2 to 38, with a mean of 15.4 (SD 7.26) and a median of 15.0. For illustrative purposes only, participants were split at the median into two equal-sized groups, respectively referred to as the Low- and High-AQ groups. Characteristics of the participants in each group are shown in Table 1, and these groups will be used in tables and figures to illustrate effects.

Final Object Clicks

Final object-click results were determined by two factors: whether the x-coordinate of the mouse when clicking the object is greater than 0, and the position of the target in that trial (such that if $x > 0$ and the target is on the right, the target has been clicked). The percentages of trials in which participants clicked on targets and distractors, for the whole study and for the High- and Low-AQ participant groups, are given in Table 2.

The binary outcome of clicking on the target versus distractor was modelled in a mixed-effects logistic regression with one within-participants predictor of manner of delivery (fluent or disfluent) and one between-participant predictor of (raw, centred) AQ score, including random intercepts and a slope for manner of delivery by participant, as well as random intercepts by target image. Participants were 2.71 times more likely to click on the target following a fluent utterance compared to a disfluent utterance ($\beta = 0.9966$, SE = 0.3052, $p = 0.0011$). There were no effects of AQ ($\beta = -0.0226$, SE = 0.0256, $p = 0.3784$), or interactions between fluency and AQ ($\beta = 0.0121$, SE = 0.0421, $p = 0.7743$).

Eye Movement

Measuring the frequencies over time with which participants fixate each of the images presented, using the eye tracking record, allows us to investigate how biases are updated in the ongoing interpretation of what is said (see Cooper, 1974; Tanenhaus et al., 1995 for more details). In the present experiment, fixation data were averaged into bins of 20 ms (10 samples). The proportions of time participants spent fixating objects were coded according to region of interest (target/distractor) for each bin. Figure 2a shows the time course of all participants' fixations to target objects and distractors over 2000 ms (100 bins) starting at target onset, for fluent and disfluent conditions respectively. For illustrative purposes, Fig. 2b, c show the equivalent patterns for the Low- and High-AQ groups of participants (Fig. 2).

We analysed eye movement data over a time window which began at target onset and ended 1500 ms later. This window was chosen based on previous research suggesting that the resolution of a referent is normally established around 400-800 ms after the object being mentioned (Eberhard et al., 1995; Hanna et al., 2003), with additional time allowed to capture any potential differences in later processing. For the purposes of analysis, we calculated a binomial "target advantage" for each bin: bins in which the majority of time was spent fixating the target were coded as 1; bins in which the majority of time was spent fixating the distractor were coded as 0. Bins in which neither target or distractor were fixated (9354) or in which both images were fixated equally (21) were discarded (10% of the data). Figure 3 shows target advantage, split by High- and Low-AQ groups, for reference.

We constructed a Bayesian generalised linear mixed model, with a logit link function, to analyse our data. We fit a model including orthogonal 1st-3rd order polynomial representations of time, allowing us to properly capture the non-linear nature of the target advantage shown in Fig. 3. As well as the time variables, the model included a within-participants and -items effect of disfluency; and a between-participants, but within-items, measure of (scaled) AQ, as well as interactions between disfluency, AQ, and each of the polynomial orders of time. Random effects for intercepts and slopes (but not interactions) were included. The model was fit using the brms package in R (Bürkner, 2017), version 2.16.1, using default priors, with four chains of 4000 iterations each. The Rhat parameter for each parameter in the model was equal to 1, indicating successful model convergence; visual inspection of posterior predictive checks suggested a good fit to the data.

Details of the converged model are given in Table 3. For each effect we report a mean effect estimate, the upper and lower bounds of a 95% credible interval, and the probability that a given effect is different from zero

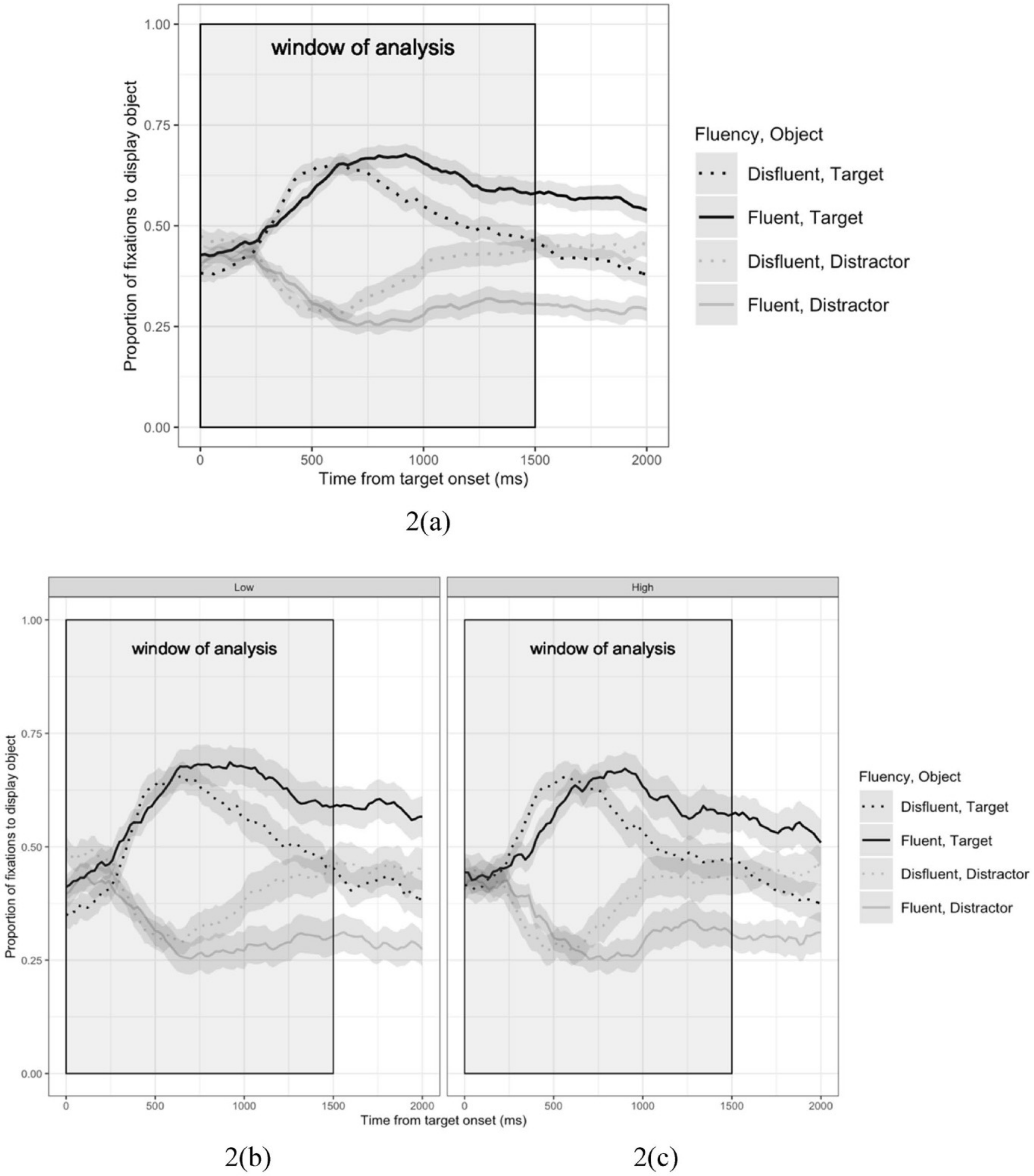


Fig. 2 Mean proportions of fixations to target object and distractor over 2000 ms from target onset, for fluent and disfluent conditions for all participants (2a), low- (2b) and high-AQ (2c) groups, calculated from summed fixations for each 20 ms time bin from target onset to 2000 ms post-onset. The shaded rectangular area represents the analysis window from target onset to 1500 ms later. The shaded areas represent ± 1 standard error of the mean by items

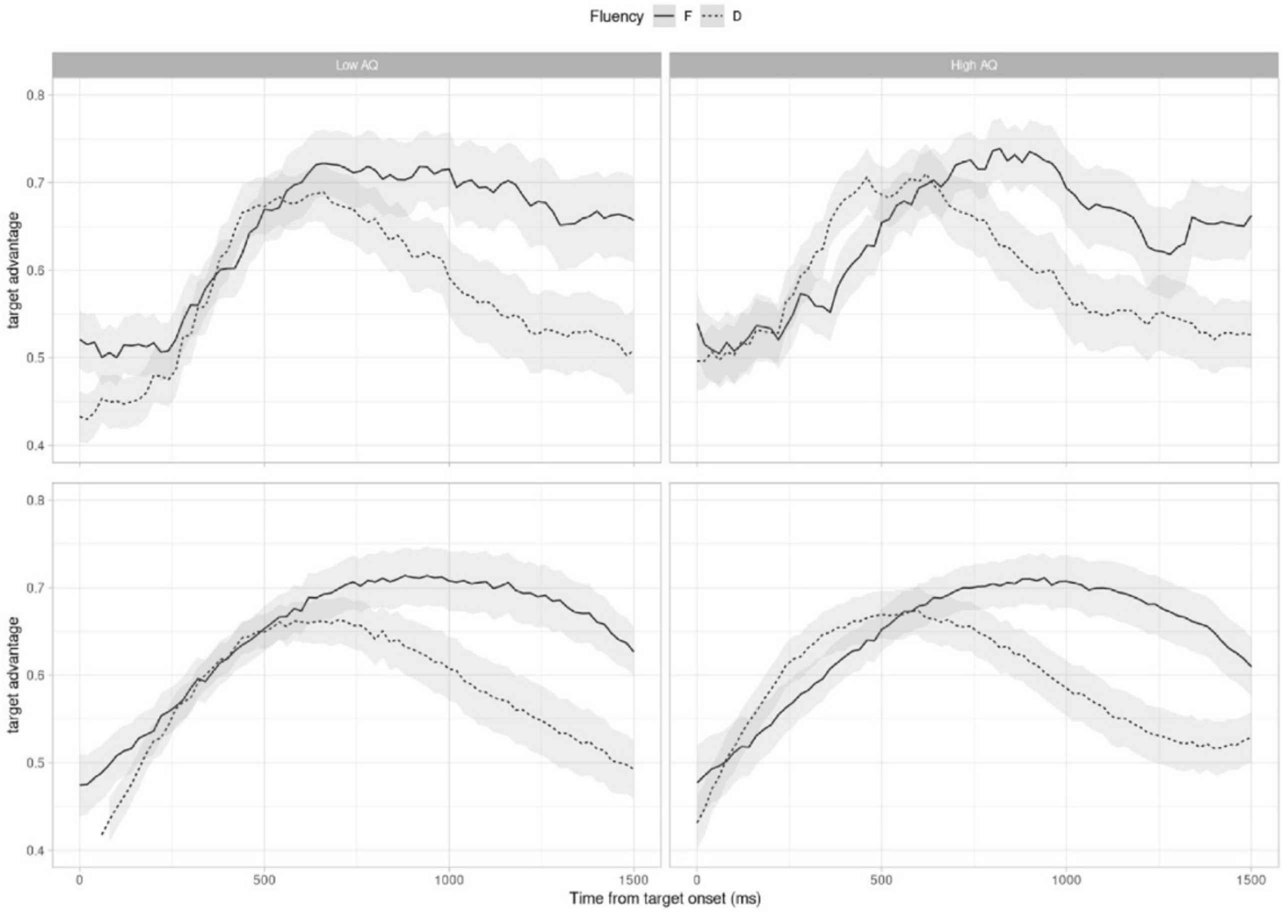


Fig. 3 Recorded (top) and model-predicted (bottom) target advantage, for Low-AQ (left) and High-AQ (right) groups of participants. Model predictions are derived from 101 posterior samples from the converged model

in the direction of its sign. There is evidence for an effect of disfluency (the probability that participants fixate the target less overall when the utterance is disfluent is 0.986) and there are notable interactions of disfluency with the polynomials of time (effectively, indicating that the curve over time of the target advantage has a different shape when utterances are disfluent). Critical to the present paper are the three-way interactions between AQ, Fluency, and each of the polynomial orders of time. There is strong evidence for each interaction (all $ps \geq 0.992$), showing that the target advantage curve changes with AQ, such that participants with higher AQ scores tend to show earlier biases towards the target. Figure 3 shows posterior predictions from the Bayesian model for the Low- and High- AQ groups, to illustrate this effect (but it is important to note that the effect is continuous, across the range of reported AQs).

Discussion

The results reported here confirm that listeners' pragmatic judgements are influenced by manner of speech, and that disfluent instructions are associated with deception. We took the greater frequency of clicks on the distractor object after hearing an instruction such as "The treasure is behind, thee, uh, hat" as evidence that participants were less inclined to believe the location of the treasure named by the speaker when the utterance was disfluent. Moreover, eye movements made during the early time-course of comprehension showed that these judgements were made quickly. These findings replicate the results of Loy et al. (2017).

Note that in these studies, participants were explicitly told that they might be deceived, and the utterances

Table 3 Model estimates for eye-tracking data, showing mean effect estimates in logits, upper and lower bounds of 95% credible intervals, and probabilities that effects differ from zero in the direction of their signs (based on posterior sampling)

Effect	b	L95%	H95%	*p* (b > / < 0)
(Intercept)	0.76	0.51	1	–
AQ	− 0.12	− 0.34	0.10	0.863
Disfluency	− 0.33	− 0.64	− 0.03	0.986
Time	2.53	1.21	3.85	1
Time^2	− 2.13	− 3.09	− 1.23	1
Time^3	− 0.54	− 1.23	0.13	0.944
AQ:Disfluency	0.02	− 0.23	0.27	0.552
AQ:Time	− 0.46	− 1.56	0.61	0.802
AQ:Time^2	0.08	− 0.55	0.71	0.590
AQ:Time^3	0.12	− 0.41	0.67	0.671
Disfluency:Time	− 2.36	− 2.63	− 2.09	1
Disfluency:Time^2	− 0.42	− 0.70	− 0.15	0.999
Disfluency:Time^3	1.36	1.09	1.63	1
AQ:Disfluency:Time	− 0.61	− 0.88	− 0.33	1
AQ:Disfluency:Time^2	0.34	0.07	0.62	0.995
AQ:Disfluency:Time^3	0.33	0.07	0.60	0.992

to which they responded were recorded. Whether there are general conclusions to be drawn about deception is unclear; here we are concerned with the processing of language that is 'known' not to have a literal interpretation. We can however draw some ecological comfort from a study by Loy et al. (2018), in which participants' judgements were affected in a similar way by disfluency in a two-player version of the current deception game, in which all utterances were spontaneous.

Critically, the present study shows that the general effects of disfluency described above vary by AQ. When hearing disfluent instructions, participants with higher AQ scores, and therefore higher numbers of traits typically associated with autism, are more likely to make early fixations on the target. The evidence for these effects is manifest as a series of interactions with a continuous AQ score predictor, and therefore not subject to criticisms typically applied to median splits (e.g., McClelland et al., 2015). However, the differences emerge as differences of processing but not of comprehension: our data do not suggest that participants' final interpretations vary with AQ, but they do suggest that there is variation in the mental processes which underlie those interpretations.

A potential account of these findings is that as the number of autistic-like traits associated with higher AQ scores increases, participants have more difficulty in utilizing contextual information. This may lead to a weaker association between disfluency and deception, attenuating the early tendency to fixate the distractor following a disfluent utterance. A variant of this account is that it may be more difficult to override more basic effects of filler disfluencies, such as their tendency to focus listeners' attention on the upcoming message (Collard et al., 2008; Fox Tree, 2002). According to the latter view, participants with higher AQs may show an initial tendency to fixate the target quickly following a disfluency, before this is overridden by slower pragmatic reasoning.

It is important to stress again that these findings are only relevant to people diagnosed with autism to the extent that AQ measures can be considered a proxy for diagnosis. The critical aim of the paper is not to identify a processing difference in the clinical population, but to point to a likely candidate which merits further investigation. Moreover, because of the continuous nature of the AQ, some statistical issues which can arise from group comparisons (especially when "controlling" for other varying factors: Miller & Chapman, 2001) are avoided. We can be confident (to the usual limits of statistical inference) that the effect is real: in the subclinical population, comprehension processes may result in the same outcomes, even though the processes by which those outcomes are achieved may vary.

In line with the study by McKenna et al. (2015), the implication of the current study is that by merely focusing on the outcomes of cognitive processing, it is possible to miss the real differences in detailed processes that vary across the population. Previous studies have used tasks in which participants are asked to provide context-appropriate explanations for non-literal speech, such as Happé's Strange Stories Test (Happé, 1994), in exploring disfluency comprehension in ASD children. However, many of these studies represent cases in which ASD- and non-ASD individuals achieved similar results, despite the ASD participants being expected to show less developed ability (Hala et al., 2007; Happé, 1995; Jolliffe & Baron-Cohen, 2000). Such results may not be conclusive without exploring the detailed cognitive processes underlying any parity in behavioural outcomes.

The present study suggests that there are differences in the processing of disfluency among individuals varying in numbers of self-identified autistic traits. This may, by extension, have implications for the comprehension processes of people with ASD diagnoses.

Author Contributions WL conceived of and designed the study, carried out initial statistical analyses, and drafted the paper; HR and MC helped with the study design, and edited the manuscript. MC did additional statistical analyses.

References

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

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning Memory and Cognition, 34*(3), 696–702. https://doi.org/10.1037/0278-7393.34.3.696

Colle, L., Baron-Cohen, S., Wheelwright, S., & Van Der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders, 38*(1), 28–40. https://doi.org/10.1007/s10803-007-0357-5

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*(6), 409–436. https://doi.org/10.1007/BF02143160

Folstein, S., & Rutter, M. (1977). Infantile autism: A genetic study of 21 twin pairs. *Journal of Child Psychology and Psychiatry, 18*(4), 297–321. https://doi.org/10.1111/j.1469-7610.1977.tb00443.x

Fox Tree, J. E. (2002). Interpreting Pauses and Ums at Turn Exchanges. *Discourse Processes, 34*(1), 37–55. https://doi.org/10.1207/s15326950dp3401_2

Gorman, K., Olson, L., Hill, A. P., Lunsford, R., Heeman, P. A., & van Santen, J. P. H. (2016). Uh and um in children with autism spectrum disorders or language impairment. *Physiology & Behavior, 9*(8), 854–865.

Hala, S., Pexman, P. M., & Glenwright, M. (2007). Priming the meaning of homographs in typically developing children and children with autism. *Journal of Autism and Developmental Disorders, 37*(2), 329–340. https://doi.org/10.1007/s10803-006-0162-6

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43–61. https://doi.org/10.1016/S0749-596X(03)00022-6

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults 1. *Journal of Autism and Developmental Disorders, 24*(2), 129–154.

Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843–855.

Heeman, P. A., Lunsford, R., Selfridge, E., Black, L., & Van Santen, J. (2010). Autism and interactional aspects of dialogue. *Proceedings of the SIGDIAL 2010 Conference: 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 249–252

Irvine, C. A., Eigsti, I. M., & Fein, D. A. (2016). Uh, um, and autism: Filler disfluencies as pragmatic markers in adolescents with optimal outcomes from autism spectrum disorder. *Journal of Autism and Developmental Disorders, 46*(3), 1061–1070. https://doi.org/10.1007/s10803-015-2651-y

Jolliffe, T., & Baron-Cohen, S. (2000). Linguistic processing in high-functioning adults with autism or Asperger's syndrome. Is global coherence impaired? *Psychological Medicine, 30*, 1169–1187.

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin and Review, 18*(1), 135–140. https://doi.org/10.3758/s13423-010-0037-x

Loukusa, S., & Moilanen, I. (2009). Pragmatic inference abilities in individuals with Asperger syndrome or high-functioning autism. A review. *Research in Autism Spectrum Disorders, 3*(4), 890–904. https://doi.org/10.1016/j.rasd.2009.05.002

Loy, J. E., Rohde, H., & Corley, M. (2016). Lying, in a manner of speaking. *Proceedings of the International Conference on Speech Prosody, 2016-Janua*, 984–988

Loy, J. E., Rohde, H., & Corley, M. (2017). Effects of disfluency in online interpretation of deception. *Cognitive Science, 41*, 1434–1456. https://doi.org/10.1111/cogs.12378

Loy, J. E., Rohde, H., & Corley, M. (2018). Cues to lying may be deceptive: Speaker and listener behaviour in an interactive game of deception. *Journal of Cognition, 1*, e42. https://doi.org/10.5334/joc.46

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McClelland, G. H., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, type II errors, and false-positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology*. https://doi.org/10.1016/j.jcps.2015.05.006

McKenna, P. E., Glass, A., Rajendran, G., & Corley, M. (2015). Strange words: Autistic traits and the processing of non-literal language. *Journal of Autism and Developmental Disorders, 45*(11), 3606–3612. https://doi.org/10.1007/s10803-015-2508-4

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*(1), 40–48. https://doi.org/10.1037/0021-843X.110.1.40

Paul, R., Orlovski, S. M., Marcinko, H. C., & Volkmar, F. (2009). Conversational behaviors in youth with high-functioning ASD and asperger syndrome. *Journal of Autism and Developmental Disorders, 39*(1), 115–125. https://doi.org/10.1007/s10803-008-0607-1

Persico, A. M., & Bourgeron, T. (2006). Searching for ways out of the autism maze: Genetic, epigenetic and environmental clues. *Trends in Neurosciences, 29*(7), 349–358. https://doi.org/10.1016/j.tins.2006.05.010

Ranick, J., Persicke, A., Tarbox, J., & Kornack, J. A. (2013). Teaching children with autism to detect and respond to deceptive statements. *Research in Autism Spectrum Disorders, 7*(4), 503–508. https://doi.org/10.1016/j.rasd.2012.12.001

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*(2), 217–236. https://doi.org/10.1068/p5117

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*(1), 25–38. https://doi.org/10.1006/jmla.1993.1002

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior, 20*, 65–80.

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior, 6*(2), 105–114. https://doi.org/10.1007/BF00987286

References

Arciuli, J., Mallard, D., & Villar, G. (2010). “Um, I can tell you're lying”: Linguistic markers of deception versus truth-telling in speech. *Applied psycholinguistics*, *31*(3), 397-411. https://doi.org/10.1017/S0142716410000044

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914. https://doi.org/10.1037/0278-7393.33.5.914

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76*(1), 28-55. https://doi.org/10.1353/lan.2000.0045.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge university press.

Bailey, K. G., & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. In *Eye movements* (pp. 487-502). Elsevier. https://doi.org/10.1016/B978-008044980-7/50024-0

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* The MIT Press.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism- Spectrum Quotient (AQ): Evidence from ... Journal of Autism and Developmental Disorders, 31(1), 5–17. Retrieved from http://arxiv.org/abs/0005074v1

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. *Oralité et gestualité: Interactions et comportements multimodaux dans la communication*, 597-600.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*(4), 441-455. https://doi.org/10.1080/01690960903047122

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and speech*, *22*(3), 201-211. https://doi.org/10.1177/002383097902200301

Beattie, I. D. (1979). Children's strategies for solving subtraction-fact combinations. *The Arithmetic Teacher*, *27*(1), 14-15. https://doi.org/10.5951/AT.27.1.0014

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1450–1460. https://doi.org/10.1037/a0027850

Blacfkmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*(3), 173-194. https://doi.org/10.1016/0010-0277(91)90052-6

Blome-Tillmann, M. (2013). Conversational implicatures (and how to spot them). *Philosophy Compass*, *8*(2), 170-185. https://doi.org/10.1111/phc3.12003

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In *The on-line study of sentence comprehension* (pp. 51-76). Psychology Press.

Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology Review*, *10*(3), 214-234. https://doi.org/10.1207/s15327957pspr1003_2

Bonnefon, J. F., Dahl, E., & Holtgraves, T. M. (2015). Some but not all dispreferred turn markers help to interpret scalar terms in polite contexts. *Thinking & Reasoning*, *21*(2), 230-249. https://doi.org/10.1080/13546783.2014.965746

Bonnefon, J. F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, *112*(2), 249-258. https://doi.org/10.1016/j.cognition.2009.05.005

Bonnefon, J. F., De Neys, W., & Feeney, A. (2011). Processing scalar inferences in face-threatening contexts. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech*, *44*(2), 123-147. https://doi.org/10.1177/00238309010440020101

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*(3), 437-457. https://doi.org/10.1016/j.jml.2004.05.006

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123-142. https://doi.org/10.1016/j.jml.2011.09.005

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423-440. https://doi.org/10.1016/j.cognition.2012.11.012

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434-463. https://doi.org/10.1016/j.cognition.2005.07.003

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of memory and language*, *44*(2), 274-296. https://doi.org/10.1006/jmla.2000.2753

Brennan, S. E., & Williams, M. (1995). The feeling of another′ s knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of memory and language*, *34*(3), 383-398. https://doi.org/10.1006/jmla.1995.1017

Brosy, J., Bangerter, A., & Mayor, E. (2016). Disfluent responses to job interview questions and what they entail. *Discourse Processes*, *53*(5-6), 371-391. https://doi.org/10.1080/0163853X.2016.1150769

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of memory and language*, *61*(2), 171-190. https://doi.org/10.1016/j.jml.2009.04.003

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, *2*(1), 11-33. https://doi.org/10.5087/dad.2011.102

Chemla, E., & Singh, R. (2014a). Remarks on the experimental turn in the study of scalar implicature, Part I. *Language and Linguistics Compass*, *8*(9), 373-386. https://doi.org/10.1111/lnc3.12081

Chemla, E., & Singh, R. (2014b). Remarks on the experimental turn in the study of scalar implicature, Part II. *Language and Linguistics Compass*, *8*(9), 387-399. https://doi.org/10.1111/lnc3.12080

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. *Structures and beyond*, *3*, 39-103.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic inquiry*, *37*(4), 535-590. 10.1162/ling.2006.37.4.535

Chomsky, N. (1965). Persistent topics in linguistic theory. *Diogenes*, *13*(51), 13-20. ttps://doi.org/10.1177/039219216501305102

Christenfeld, N. (1994). Options and ums. *Journal of Language and Social Psychology*, *13*(2), 192-199. https://doi.org/10.1177/0261927X94132005

Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73-111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clark, H. H. (1999). Passing time in conversation. In *workshop on the structure of spoken and written texts, University of Texas, Austin*.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. Cognition, 84(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention Orienting Effects of Hesitations in Speech: Evidence From ERPs. Journal of Experimental Psychology: Learning Memory and Cognition, 34(3), 696–702. https://doi.org/10.1037/0278-7393.34.3.696

Colle, L., Baron-Cohen, S., Wheelwright, S., & Van Der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. Journal of Autism and Developmental Disorders, 38(1), 28–40. https://doi.org/10.1007/s10803- 007-0357-5

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, No. 31).

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. Cognitive Psychology, 6(1), 84-107. https://doi.org/10.1016/0010-0285(74)90005-X

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass*, *2*(4), 589-602. https://doi.org/10.1111/j.1749-818X.2008.00068.x

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658-668. https://doi.org/10.1016/j.cognition.2006.10.010

Cossavella, F., & Cevasco, J. (2021). The importance of studying the role of filled pauses in the construction of a coherent representation of spontaneous spoken discourse. *Journal of Cognitive Psychology*, *33*(2), 172-186. https://doi.org/10.1080/20445911.2021.1893325

Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis) fluency. *Languages in Contrast*, *17*(1), 69-95. https://doi.org/10.1075/lic.17.1.04cri

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & cognition*, *35*, 15-28. https://doi.org/10.3758/BF03195938

de Boer, M. M., & Heeren, W. F. (2020). Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English. *The Journal of the Acoustical Society of America*, *148*(6), 3612-3622. https://doi.org/10.1121/10.0002871

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly*, *15*(3), 237-254. https://doi.org/10.1080/15434303.2018.1477780

De Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics*, *19*(2), 85-114. https://doi.org/10.1017/S1470542707000049

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*, 1-12. https://doi.org/10.3758/s13428-014-0458-y

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, *39*(4), 667-710. https://doi.org/10.1111/cogs.12171

DePaulo, B. M., Rosenthal, R., Rosenkrantz, J., & Green, C. R. (1982). Actual and perceived cues to deception: A closer look at speech. *Basic and Applied Social Psychology*, *3*(4), 291-312. https://doi.org/10.1207/s15324834basp0304_6

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. Journal of Psycholinguistic Research, 24(6), 409–436. https://doi.org/10.1007/BF02143160

Engelhardt, P. E., McMullon, M. E., & Corley, M. (2019). Individual differences in the production of disfluency: A latent variable analysis of memory ability and verbal intelligence. *Quarterly journal of experimental psychology*, *72*(5), 1084-1101. https://doi.org/10.1177/1747021818778752

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2013). Is the fluency of language outputs related to individual differences in intelligence and executive function?. *Acta psychologica*, *144*(2), 424-432. https://doi.org/10.1016/j.actpsy.2013.08.002

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive science*, *31*(5), 889-909. https://doi.org/10.1080/03640210701530797

Feeney, A., & Bonnefon, J. F. (2013). Politeness and honesty contribute additively to the interpretation of scalar expressions. *Journal of Language and Social Psychology*, *32*(2), 181-190. https://doi.org/10.1177/0261927X12456840

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57-84. https://doi.org/10.1006/jmla.2001.2797

Ferreira, F., Lau, E. F., & Bailey, K. G. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, *28*(5), 721-749. https://doi.org/10.1207/s15516709cog2805_5

Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker?. *Psychonomic bulletin & review*, *19*, 921-928. https://doi.org/10.3758/s13423-012-0279-x

Folstein, S., & Rutter, M. (1977). Infantile Autism: a Genetic Study of 21 Twin Pairs. Journal of Child Psychology and Psychiatry, 18(4), 297–321. https://doi.org/10.1111/j.1469- 7610.1977.tb00443.x

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of memory and language*, *34*(6), 709-738. https://doi.org/10.1006/jmla.1995.1032

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & cognition*, *29*, 320-326. https://doi.org/10.3758/BF03194926

Fox Tree, J. E. (2002). Interpreting Pauses and Ums at Turn Exchanges. Discourse Processes, 34(1), 37–55. https://doi.org/10.1207/s15326950dp3401_2

Francesch, P., & Payrató, L. (2024). Pragmatic Ambiguity, Implicatures, and Translation 1. *Studias Linguistica*, *78*(1), 156-185. https://doi.org/10.1111/stul.12219

Fraser, B. (2010). Pragmatic competence: The case of hedging. In *New approaches to hedging* (pp. 15-34). Brill. https://doi.org/10.1163/9789004253247_003

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178-210. https://doi.org/10.1016/0010-0285(82)90008-1

Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519-559. https://doi.org/10.1007/BF00138988

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior research methods*, *42*(1), 226-241. https://doi.org/10.3758/BRM.42.1.226

Garten, J., Kennedy, B., Sagae, K., & Dehghani, M. (2019). Measuring the importance of context when modeling language comprehension. *Behavior research methods*, *51*, 480-492. https://doi.org/10.3758/s13428-019-01200-w

Goffman, E. (1967). *Interaction ritual: essays on face-to-face interaction.* Aldine.

Goldman-Eisler, F. (1961). The significance of changes in the rate of articulation. *Language and Speech*, *4*(3), 171-174. https://doi.org/10.1177/002383096100400305

Goodale, M. A., Pelisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, *320*(6064), 748-750. https://doi.org/10.1038/320748a0

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological inquiry*, *50*(3-4), 272-302.

Gorman, K., Olson, L., Hill, A. P., Lunsford, R., Heeman, P. A., & Santen, J. P. H. van. (2016). Uh and um in Children With Autism Spectrum Disorders or Language Impairment. Physiology & Behavior, 9(8), 854–865.

Gósy, M. (2001). The double function of disfluency phenomena in spontaneous speech. In *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41-58). Brill. https://doi.org/10.1163/9789004368811_003

Grodner, D. J., & Sedivy, J. (2011). The effect of speaker-specific information on pragmatic inferences. https://doi.org/10.7551/mitpress/8957.003.0013

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42-55. https://doi.org/10.1016/j.cognition.2010.03.014

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *science*, *304*(5669), 438-441. 10.1126/science.1095455

Hala, S., Pexman, P. M., & Glenwright, M. (2007). Priming the meaning of homographs in typically developing children and children with autism. Journal of Autism and Developmental Disorders, 37(2), 329–340. https://doi.org/10.1007/s10803-006-0162-6

Hallin, A. E., Garcia, G. D., & Reuterskiöld, C. (2016). The use of causal language and filled pauses in children with and without autism. *Child Development Research*, *2016*(1), 8535868. https://doi.org/10.1155/2016/8535868

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. Journal of Memory and Language, 49(1), 43–61. https://doi.org/10.1016/S0749-596X(03)00022-6

Happé, F. G. E. (1994). An Advanced Test of Theory of Mind: Understanding of Story Characters' Thoughts and Feelings by Able Autistic, Mentally Handicapped, and Normal Children and Adults 1. Journal of Autism and Developmental Disorders, 24(2), 129–154.

Happé, F. G. E. (1995). The Role of Age and Verbal Ability in the Theory of Mind Task Performance of Subjects with Autism Author (s): Francesca G . E . Happé Published by: Wiley on behalf of the Society for Research in Child Development Stable URL : http://www.jstor.org/stabl. Child Development, 66(3), 843–855.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of educational psychology*, *56*(4), 208. https://doi.org/10.1037/h0022263

Heeman, P. A., Lunsford, R., Selfridge, E., Black, L., & Van Santen, J. (2010). Autism and interactional aspects of dialogue. Proceedings of the SIGDIAL 2010 Conference: 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 249–252.

Hendrickson, K., Apfelbaum, K., Goodwin, C., Blomquist, C., Klein, K., & McMurray, B. (2022). The profile of real-time competition in spoken and written word recognition: More similar than different. *Quarterly Journal of Experimental Psychology*, *75*(9), 1653-1673. https://doi.org/10.1177/17470218211056842

Holtgraves, T., & Kraus, B. (2018). Processing scalar implicatures in conversational contexts: An ERP study. *Journal of Neurolinguistics*, *46*, 93-108. https://doi.org/10.1016/j.jneuroling.2017.12.008

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. University of California, Los Angeles.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive psychology*, *58*(3), 376-415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-B32. https://doi.org/10.1016/j.cognition.2004.10.003

Irvine, C. A., Eigsti, I. M., & Fein, D. A. (2016). Uh, Um, and Autism: Filler Disfluencies as Pragmatic Markers in Adolescents with Optimal Outcomes from Autism Spectrum Disorder. Journal of Autism and Developmental Disorders, 46(3), 1061–1070. https://doi.org/10.1007/s10803-015-2651-y

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461-3493. https://doi.org/10.3758/s13428-022-01969-3

Jefferson, G. (1974). Error correction as an interactional resource1. *Language in society*, *3*(2), 181-199.

Jolliffe, T., & Baron-Cohen, S. (2000). Linguistic processing in high-functioning adults with autism or Asperger's syndrome. Is global coherence impaired? (October 2000), 1169–1187.

Katsos, N. (2008). The semantics/pragmatics interface from an experimental perspective: the case of scalar implicature. *Synthese*, *165*(3), 385-401. https://doi.org/10.1007/s11229-007-9187-4

Kennedy, C. (2011). Ambiguity and vagueness: An overview. *Semantics: An international handbook of natural language meaning*, *1*, 507-535.

King, J. P., Loy, J. E., & Corley, M. (2018). Contextual effects on online pragmatic inferences of deception. *Discourse Processes*, *55*(2), 123-135. https://doi.org/10.1080/0163853X.2017.1330041

King, J. P., Loy, J. E., Rohde, H., & Corley, M. (2020). Interpreting nonverbal cues to deception in real time. *PLoS One*, *15*(3), e0229486. https://doi.org/10.1371/journal.pone.0229486

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145-164. https://doi.org/10.1016/j.system.2004.01.001

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*(2), 335-342. https://doi.org/10.1016/j.cognition.2014.05.017

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. Psychonomic Bulletin and Review, 18(1), 135–140. https://doi.org/10.3758/s13423- 010-0037-x

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* The MIT Press.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Lickley, R. (2017). Disfluency in typical and stuttered speech. *Book series Studi AISV*, *3*, 373-387. https://doi.org/10.17469/O2103AISV000019

Lickley, R. J. (2015). Fluency and disfluency. *The handbook of speech production*, 445-474. https://doi.org/10.1002/9781118584156.ch20

Loukusa, S., & Moilanen, I. (2009). Pragmatic inference abilities in individuals with Asperger syndrome or high-functioning autism. A review. Research in Autism Spectrum Disorders, 3(4), 890–904. https://doi.org/10.1016/j.rasd.2009.05.002

Loy, J. E., Rohde, H., & Corley, M. (2016). Lying, in a manner of speaking. In *Proceedings of speech prosody* (Vol. 8, p. 3). https://doi.org/10.21437/speechprosody.2016-202

Loy, J. E., Rohde, H., & Corley, M. (2017). Effects of Disfluency in Online Interpretation of Deception. Cognitive Science, 41, 1434–1456. https://doi.org/10.1111/cogs.12378

Loy, J. E., Rohde, H., & Corley, M. (2018). Cues to lying may be deceptive: Speaker and listener behaviour in an interactive game of deception. Journal of Cognition, 1, e42. https://doi.org/10.5334/joc.46

Loy, J. E., Rohde, H., & Corley, M. (2019). Real-time social reasoning: the effect of disfluency on the meaning of some. *Journal of Cultural Cognitive Science*, *3*(2), 159-173. https://doi.org/10.1007/s41809-019-00037-1

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of memory and language*, *32*(5), 692-715. https://doi.org/10.1006/jmla.1993.1035

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, *101*(4), 676. https://doi.org/10.1037/0033-295X.101.4.676

MacGregor, L. J. (2008). Disfluencies affect language comprehension: evidence from event-related potentials and recognition memory. (Doctoral dissertation, University of Edinburgh)

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19-44. https://doi.org/10.1080/00437956.1959.11659682

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202. https://doi.org/10.1037/0096-3445.132.2.202

Marslen-Wilson, W. D., & Tyler, L. K. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *295*(1077), 317-332. https://doi.org/10.1098/rstb.1981.0143

Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception & Psychophysics*, *3*(6), 427-438. https://doi.org/10.3758/BF03205750

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods, 44(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McClelland, G. H., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, Type II errors, and false-positive consumer psychology: Don't fight the power. Journal of Consumer Psychology. https://doi.org/10.1016/j.jcps.2015.05.006

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and cognitive processes*, *4*(3-4), SI287-SI335. https://doi.org/10.1080/01690968908406371

McKenna, P. E., Glass, A., Rajendran, G., & Corley, M. (2015). Strange Words: Autistic Traits and the Processing of Non-Literal Language. Journal of Autism and Developmental Disorders, 45(11), 3606–3612. https://doi.org/10.1007/s10803-015-2508-4

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22-24. https://doi.org/10.1111/j.1467-9280.2008.02041.x

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. Journal of Abnormal Psychology, 110(1), 40–48. https://doi.org/10.1037/0021-843X.110.1.40

Moniz, H. G. S. (2013). *Processing disfluencies in european portuguese* (Doctoral dissertation, Universidade de Lisboa (Portugal)).

Nolte, C., Seedorff, M., Oleson, J., Brown, G., Cavanaugh, J., & McMurray, B. (2023). *bdots: Bootstrapped Differences of Time Series* (version 1.2.5). https://cran.r-project.org/web/packages/bdots/index.html

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and language*, *85*(2), 203-210. https://doi.org/10.1016/S0093-934X(03)00053-1

Novick, J. M., Kim, A., & Trueswell, J. C. (2003). Studying the grammatical aspects of word recognition: Lexical priming, parsing, and syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, *32*, 57-75. https://doi.org/10.1023/A:1021985032200

Oleson, J. J., Cavanaugh, J. E., McMurray, B., & Brown, G. (2017). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical methods in medical research*, *26*(6), 2708-2725. https://doi.org/10.1177/0962280215607411

Oomen, C. C., & Postma, A. (2001). Effects of divided attention on the production of filled pauses and repetitions. https://doi.org/10.1044/1092-4388(2001/078)

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, *9*(1), 19-36.

Panizza, D., Chierchia, G., & Clifton Jr, C. (2009). On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of memory and language*, *61*(4), 503-518. https://doi.org/10.1016/j.jml.2009.07.005

Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software, 80(1), 1-28. http://doi.org/10.18637/jss.v080.i01

Paul, R., Orlovski, S. M., Marcinko, H. C., & Volkmar, F. (2009). Conversational behaviors in youth with high-functioning ASD and asperger syndrome. Journal of Autism and Developmental Disorders, 39(1), 115–125. https://doi.org/10.1007/s10803-008-0607-1

Pehar, D. (2001). Use of ambiguities in peace agreements. *Language and diplomacy*, 163-200.

Persico, A. M., & Bourgeron, T. (2006). Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends in neurosciences*, *29*(7), 349-358. https://doi.org/10.1016/j.tins.2006.05.010

Politzer-Ahles, S., & Fiorentino, R. (2013). The realization of scalar inferences: Context sensitivity without processing cost. *PloS one*, *8*(5), e63943. https://doi.org/10.1371/journal.pone.0063943

Politzer-Ahles, S., & Husband, E. M. (2018). Eye movement evidence for context-sensitive derivation of scalar inferences. *Collabra: Psychology*, *4*(1), 3. https://doi.org/10.1525/collabra.100

Postma, A., & Kolk, H. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech, Language, and Hearing Research*, *35*(3), 537-544. https://doi.org/10.1044/jshr.3503.537

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, *39*(4), 375-392. https://doi.org/10.1177/002383099603900403

Postma, A., Kolk, H., & Povel, D. J. (1990). Speech planning and execution in stutterers. *Journal of Fluency Disorders*, *15*(1), 49-59. https://doi.org/10.1016/0094-730X(90)90032-N

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Ranick, J., Persicke, A., Tarbox, J., & Kornack, J. A. (2013). Teaching children with autism to detect and respond to deceptive statements. Research in Autism Spectrum Disorders, 7(4), 503–508. https://doi.org/10.1016/j.rasd.2012.12.001

Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, *2*, 51-81. https://doi.org/10.1007/BF01067111

Rose, R. L. (2008). *Filled pauses in language teaching: Why and how* (Doctoral dissertation, Waseda University).

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. Perception, 33(2), 217–236. https://doi.org/10.1068/p5117

Sarrett, M. E., Shea, C., & McMurray, B. (2022). Within-and between-language competition in adult second language learners: implications for language proficiency. *Language, Cognition and Neuroscience*, *37*(2), 165-181. https://doi.org/10.1080/23273798.2021.1952283

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of personality and social psychology*, *60*(3), 362. https://doi.org/10.1037/0022-3514.60.3.362

Schachter, S., Rauscher, F., Christenfeld, N., & Crone, K. T. (1994). The vocabularies of academia. *Psychological Science*, *5*(1), 37-41. https://doi.org/10.1111/j.1467-9280.1994.tb00611.x

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 28, No. 28).

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more). *Journal of memory and language*, *102*, 55-67. https://doi.org/10.1016/j.jml.2018.05.004

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies* (Doctoral dissertation, University of California, Berkeley).

Shriberg, E. (2001). To 'errrr'is human: ecology and acoustics of speech disfluencies. *Journal of the international phonetic association*, *31*(1), 153-169. 10.1017/S0025100301001128

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. Journal of memory and language, 32(1), 25-38. https://doi.org/10.1006/jmla.1993.1002

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological science*, *10*(3), 281-284. https://doi.org/10.1111/1467-9280.00151

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1521. https://doi.org/10.1037/0278-7393.24.6.1521

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227-267. https://doi.org/10.1016/0010-0277(94)00647-4

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *47*(2), 276. https://doi.org/10.1037/h0078826

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393-10398. https://doi.org/10.1073/pnas.0503903102

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive psychology*, *45*(4), 447-481. https://doi.org/10.1016/S0010-0285(02)00503-0

Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language*, *86*(2), 326-343. https://doi.org/10.1016/S0093-934X(02)00547-3

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. https://doi.org/10.7551/mitpress/1479.003.0029

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In *Handbook of psycholinguistics* (pp. 863-900). Academic Press. https://doi.org/10.1016/B978-012369374-7/50023-7

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268(5217), 1632-1634. http://doi.org/10.1126/science.7777863

Tomlinson Jr, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of memory and language*, *69*(1), 18-35. https://doi.org/10.1016/j.jml.2013.02.003

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of memory and language*, *35*(4), 566-585. https://doi.org/10.1006/jmla.1996.0030

Van Berkum, J. J. (2009). The neuropragmatics of'simple'utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory* (pp. 276-316). Palgrave Macmillan. https://hdl.handle.net/11858/00-001M-0000-0013-1663-F

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of cognitive neuroscience*, *20*(4), 580-591. 10.1162/jocn.2008.20054

Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. Journal of Nonverbal Behavior, 20, 65–80.

Walker, E. J., Risko, E. F., & Kingstone, A. (2014). Fillers as signals: Evidence from a question–answering paradigm. *Discourse Processes*, *51*(3), 264-286. https://doi.org/10.1080/0163853X.2013.862478

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech communication*, *50*(2), 81-94. https://doi.org/10.1016/j.specom.2007.06.002

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86-116. https://doi.org/10.1016/j.wocn.2018.03.002

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*(1), 3-36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Yuan, J., Xu, X., Lai, W., & Liberman, M. (2016, May). Pauses and pause fillers in Mandarin monologue speech: The effects of sex and proficiency. In *Proceedings of Speech Prosody* (Vol. 2016, pp. 1167-1170).

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In *Advances in experimental social psychology* (Vol. 14, pp. 1-59). Academic Press. https://doi.org/10.1016/S0065-2601(08)60369-X

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. Journal of Nonverbal Behavior, 6(2), 105–114. https://doi.org/10.1007/BF00987286